2-1-2013 12:00 AM

Functional characterization of the HUA2 gene family in Arabidopsis thaliana

Preetam Janakirama
The University of Western Ontario

Supervisor
Dr. Vojislava Grbic
The University of Western Ontario

Graduate Program in Biology
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

Functional characterization of the *HUA2* gene family in *Arabidopsis thaliana*

(Thesis format: Monograph)

by

Preetam Janakirama

Graduate Program in Biology

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Supervisor

_____
Dr. Vojislava Grbic

Examiners

_____
Dr. Sashko Damjanovki

_____
Dr. Robert Cumming

_____
Dr. Lars Konermann

_____
Dr. Thomas Berleth

The thesis by

Preetam Janakirama

entitled:

Functional characterization of the *HUA2* gene family in *Arabidopsis thaliana*

is accepted in partial fulfilment of the
requirements for the degree of Doctor of Philosophy

Date_____

_____
Chair of the Thesis Examination Board

Abstract

HUA2 encodes a key developmental regulatory protein implicated in the coordination of induction and the maintenance of floral state in *Arabidopsis*. To gain further insight into the function of HUA2, I have conducted a series of studies aimed at elucidating the molecular function(s) of its individual domains. I show that the PWWP, RPR and CT-proline rich domains within HUA2 are required for the proper regulation of the flowering time phenotype. I also establish that HUA2 interacts with characterized splicing factors (FCA, AtPRP40, RBP45 and UBP1) through its CT-proline rich domain. In addition, I examine the overlap in function between HUA2 and its related family members (the *HUA2*-like genes, *HULK1, HULK2* and *HULK3*). I show that all members of the family localize to the nucleus where they interact with the same set of pre-mRNA processing factors (FCA, AtPRP40, RBP45 and UBP1). Lastly, I show that HUA2 interacts with the AtCTD via binding to CTD heptad repeats that are phosphorylated exclusively on Ser-2 residues. This key result is consistent with my data demonstrating the interaction of HUA2 with splicing factors, and strongly suggests a role for HUA2 in the 3'-end processing of pre-mRNAs. The possible involvement of HUA2 in regulating the pre-mRNA processing of genes involved in floral development is discussed.

Key words: *HUA2, HULK1, HULK2, HULK3, FLC, AtCTD*, pre-mRNA processing, PWWP domain and RPR domain.

Acknowledgements

It is a great pleasure to thank everyone who helped me during my Ph.D. program and those who have helped me write my dissertation. I owe sincere and earnest thankfulness to my supervisor, Dr. Vojislava Grbic, for giving me the opportunity to do my Ph.D. in her laboratory. I greatly appreciate the support and guidance she provided me throughout my program and thesis writing. I would like to acknowledge my advisory committee members, Dr. Jim Karagiannis and Dr. Priti Krishna, for their valuable advice and input used to develop this research. I am truly indebted and thankful to Dr. Jim Karagiannis for his thorough and critical review of my dissertation as my Reader.

I thank Dr. David Darby and Miss Verena Bremer for their support in helping me obtain an Ontario/Baden-Württemberg (OBW) program scholarship, which gave me an opportunity to carry out some of my research experiments at the Center for Plant Molecular Biology (ZMBP), Eberhard Karls University of Tübingen, Germany. I am thankful and grateful to Dr. Sascha Laubinger, his laboratory and the entire department in ZMBP for all of their help with my research during my stay in Germany. Thank you to Dr. Robert Cumming and his laboratory members for their kindness and support with western blot analysis, Dr. Rima Menassa and Dr. Igor Kolotin at Agriculture and Agri-Food Canada for helping with particle bombardment experiments and Dr. M. K. Mathew at the National Centre for Biological Sciences (NCBS), Bangalore, India and Dr. K. Muniyappa at the Indian Institute of Science (IISc), Bangalore, India for giving me an opportunity to work in their lab, which provided the foundation for my research work before coming to Western. I would also like to thank Dr. Robert Dean, my TA supervisor, for being a positive role model and an excellent teaching mentor and Patricia gray for her kindness and for providing me teaching assistant positions till the end of my program.

I thank all the present and past members of my laboratory who have helped and contributed to my research project. Many thanks to Uday Sajja, Sathya Jali and Sarah Rosloski for their guidance in the laboratory towards my project; especially to Sarah, who provided insightful feedback on my thesis. I want to thank Vladimir Zhurov who has been like an elder brother and provided me with a lot of help and guidance in the laboratory. Thank you to Biljana Popovic for assistance with laboratory procedures.

Thanks to Scott Deroo for preparing some of the constructs I used in my work. Thanks to all of the volunteers, work-study students and research assistants, especially Akosh Kazinczi.

I could not have completed my Ph.D. without the guidance of my father, who is a constant source of inspiration and encouragement in my life. Thank you to all of my family back in India for their moral support. I am very grateful to Holly Stover for her tremendous help and generosity, especially during difficult periods, and for always believing in me. This thesis would not have been possible without her assistance. I also want to thank my friends at Western for all of their support and kindness including Dan Bath, Christopher Doan, Scott Deroo, Mathias Sinnocca, Cherise Ens, Ingrid Fung, Tara Negrave, Eridan Pereira, Lovesha Sivanantharajah, Alissa Bults, Jordan Epstein, Anita Prtenjaca, Andrea Wishart, Rainer Bode, Robert Arseneault, Mark Fox, Shyamsundar Kulkarni, Srikanth Talluri and Vishal Gupta.

Last but not least, I would like to thank my friends back in India Girish Anil Kumar, Santosh B Gowda and Goutam Narasimhan for their moral support. And to my dear friend Mukund Ramkrishnan, without him I would have never made it to graduate school nor would have survived. His will, honesty and perseverance has and always will inspire me in my life.

Table of contents

List of Tables

List of figures

List of Appendices

List of Abbreviations

°C	degree Celsius
µg	microgram
µl	microliter
aa	amino acid
ATP	adenosine triphosphate
bp	base pair
BSA	bovine serum albumin
CaMV	cauliflower mosaic virus
cDNA	complementary deoxyribonucleic acid
Col-0	Columbia, accession of *Arabidopsis*
Cm	centimeter
CTD	c-terminal domain of RNA polymerase II
d	day
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FWD	Forward
g	Grams
GA	gibberellin pathway
h	hour
JGI	joint genome institute

kb	kilo base pair
kV	kilo volts
L	Liter
LB	Laura Bretani medium
LD	long days
Mb	megabase
MCS	multiple cloning site
Min	minute
MOPS	3-[N-morpholino]propanesulfonic acid
mRNA	messenger ribonucleic acid
½ MS	half-strength murashige and skoog medium
ml	mili liters
ms	milli second
nt	nucleotides
OD_{600}	optical density at 600
ORF	open reading frame
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PHD	plant homeodomain
PPLP	proline-proline-lysine-proline
PRC2	polycomb repression complex 2
Psi	per square inch
PWWP	proline-tryptophan-tryptophan-proline
REV	Reverse

RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
RPR	regulation of nuclear pre-mRNA processing
RT	room temperature
SAM	shoot apical meristem
SD	short days
snRNPs	small ribonucleoproteins
Std	standard deviation
SDS	sodium dodecyl sulphate
SSC	saline-sodium citrate
T-DNA	transfer-deoxyribonucleic acid
UTR	untranslated region
WW	tryptophan-tryptopha

Chapter 1: Introduction

1.1 *Arabidopsis thaliana* as a model system

Arabidopsis thaliana is a small dicotyledonous plant species that belongs to the mustard family (*Brassicaceae*). It is also known as mouse-ear cress and is considered a weed, even though it is closely related to economically important crops like turnip, broccoli, canola and cabbage. The entire life cycle of *A.thaliana* is completed in approximately six weeks. The seedlings first develop a rosette of leaves approximately 2–10 cm in diameter (Figure. 1.1). Bolting – the development of a stalk with flowers – begins at the transition to reproductive development and occurs anywhere from 2–3 weeks to several months after germination. The fruits of the plants or siliques (Figure 1.1) can cumulatively produce up to 10,000 seeds.

A. thaliania has been one of the most important model organisms for plant biology for several decades. It has been used in the study of plant physiology, genetics and molecular biology. It requires very limited space and can be easily grown in an indoor growth chamber or greenhouse. In addition, the short generation time of the species in laboratory, reproduction primarily by self-fertilization, rapid flowering, production of a large number of seeds and a small genome make it an ideal model species (Dean, 1993; Meyerowitz and Somerville, 1994; Meinke et al., 1998; Pyke, 1994). These advantages have inspired the plant research community to investigate the biological processes of *Arabidopsis* and has led them to characterize many genes (Meinke et al., 1998). A key milestone in *Arabidopsis* research was the publication of its annotated genome sequence.

Figure 1.1. *Arabidopsis thaliana*. *Arabidopsis* is a small weed belonging to the Brassicaceae (mustard family). Floral development begins about three weeks after germination in early flowering strains and mature plants (A) reach 15–20 cm in height. *Arabidopsis* can produce up to 500 seeds per silique (fruit) (B). The cauline leaves (C) of the plant begin development approximately 2.5 weeks after germination. The basal rosette (D) is approximately 2–10 cm in diameter.

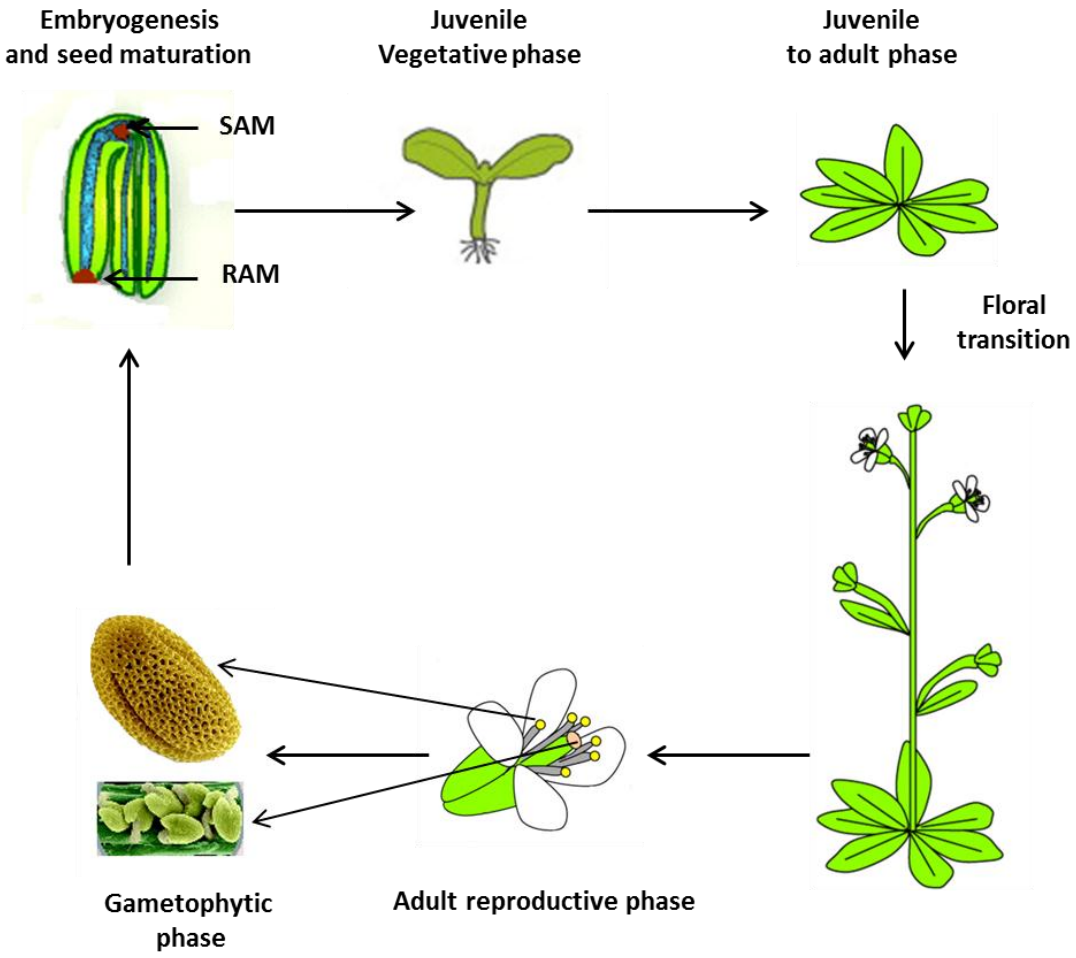

A
B
C
D

According to The Arabidopsis Information Resource (TAIR), *Arabidopsis* contains 27,416 protein coding genes, 4,827 pseudogenes or transposable element genes and 1,359 ncRNA. In total there are 33,597 genes and 41,666 gene models (ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/README_TTAI10.txt, on www.arabidopsis.org, Nov 17 2010). Less than 10% of *Arabidopsis* genes have been investigated and functionally characterized. The availability of the annotated genome sequence thus provides the foundation for further detailed functional characterization of plant genes. Out of all plant species used as model organisms, *Arabidopsis* offers many advantages and was thus chosen as the model organism for my research project.

1.2 Different stages in plant development

The angiosperm life cycle is characterized by transition through unique developmental growth phases: germination, a vegetative phase (with a juvenile and adult component), reproductive flowering phase, seed set and senescence (Figure 1.2) (Huijser and Schmid, 2011). Development through each unique growth phase is controlled by genetic programs that are strongly influenced by, and are dependent upon, environmental and internal signalling. The majority of the plant's growth occurs during the vegetative stage when photosynthetic capacity, size and mass increase. The juvenile and adult growth forms are usually highly distinct but the differences are more pronounced in perennials than annuals and species such as *Arabidopsis* have very subtle changes. The transition from the juvenile to the adult vegetative growth phase consists of simultaneous changes in leaf

Figure 1.2. Different phases of plant development. After fertilization, shoot apical meristem (SAM) and root apical meristems (RAM) are established in the developing embryo. These meristems are the source of all the post-embryonic organs formed throughout the life of the plant. Following germination, the plant enters a vegetative phase, which begins with a juvenile period and ends with reproductive competence and flowering. Under suitable conditions, the plant switches from vegetative to reproductive phase when the SAM produces bisexual flowers instead of shoots. The plant enters the haploid gametophytic phase when the flower produces both male and female gametes. The gametes fuse to form a new diploid zygote which develops into an embryo. Figure adapted and modified from (Huijser and Schmid, 2011)


Embryogenesis
and seed maturation
Juvenile
Vegetative phase
Juvenile
to adult phase
SAM
RAM
Floral
transition
Gametophytic
phase
Adult reproductive phase

size and shape, trichome distribution and internode length and normally requires reproductive competence. Therefore, reproductive organs are usually only formed by adult plants. Sexual reproduction occurs after plants become competent to flower and undergo reproductive development. The gametophytic phase of sexual reproduction is initiated by the production of gametes. Male and female gametes are formed within the developing flower and a short gametophytic haploid development phase takes place. After this the male and female gametes combine to form a new diploid zygote which initiates the sporophytic phase. Actively dividing stem cells of developing embryos form two distinct populations at opposite ends of the primary growth axis which are called the root apical meristem (RAM) and the shoot apical meristem (SAM). All post-embryonic growth during the plant's life cycle originates from the meristems and all above ground organs (shoots and flowers) form from the SAM. During reproductive development, the SAM only produces flowers and shoot development does not normally occur (Huijser and Schmid, 2011).

1.3 Transition from vegetative to reproductive development in plants

Flowering is a critical process in the life cycle of any angiosperm. This event is highly sensitive to environmental and internal triggers. Only when the environmental conditions are favorable for reproduction will the plant switch from the vegetative growth stage to the reproductive stage. Understanding this biological transition has been one of the significant challenges in modern plant biology research. The ability of the plant to perceive and respond to environmental cues is correlated with the survival of the species in a specific habitat (Willis et al., 2008). It is therefore no surprise that plants have developed a finely regulated network to integrate diverse inputs from environmental and

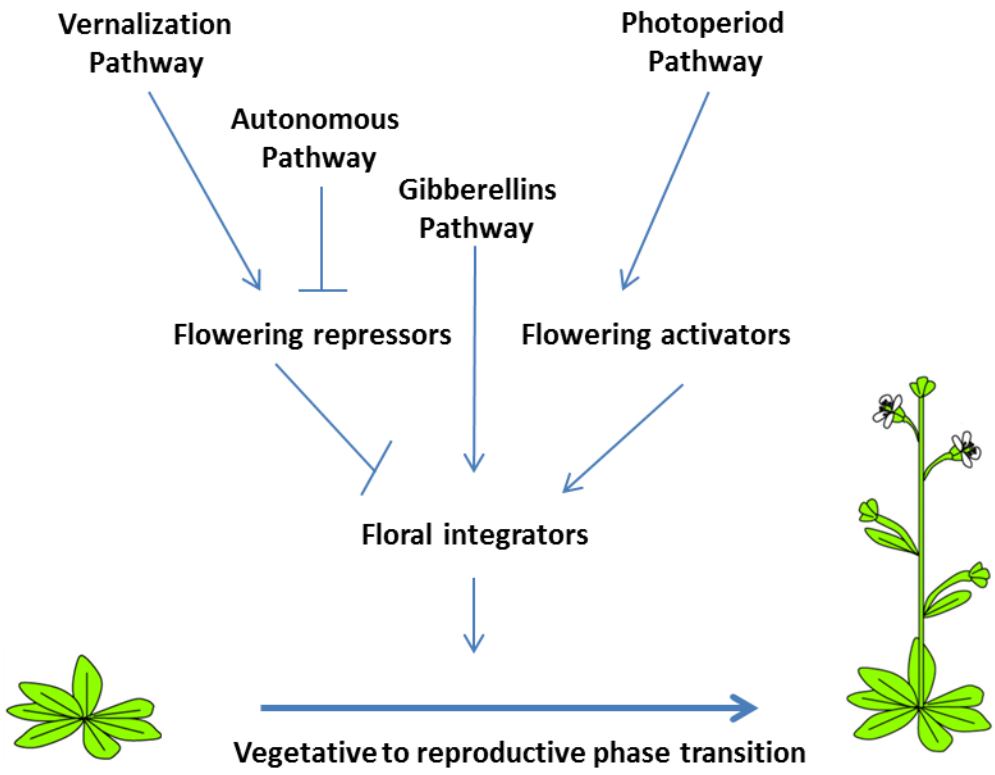
endogenous signals and use their resources appropriately to initiate flowering when the conditions are most favorable (Yant et al., 2010). The regulatory network that controls the reproductive transition is achieved by a fine balance between activators and repressors, both of which play important roles at every stage of flowering network and operate throughout plant tissue (Kobayashi and Weigel, 2007; Giakountis and Coupland, 2008; Yant et al., 2010).

Extensive research on genetic and molecular studies in *Arabidopsis* has identified four major flowering pathways; the photoperiod, vernalization, autonomous and gibberelic acid pathways (Figure 1.3). Each of these pathways contributes to floral promotion, floral inhibition, or to a neutral state depending on the plant's response to different cues (Yant et al., 2010). One of the major environmental signals regulating flowering is the photoperiod pathway. The photoperiod pathway promotes flowering in response to day and night cycles. Photoperiod is sensed in the leaves and promotes changes in the SAM. *FLOWERING LOCUS T (FT)* is a major regulator of transition to flowering by photoperiod inputs. It is expressed in leaves, and the protein is later transported to the shoot apex, making the SAM switch from leaf production to early floral development (Jaeger et al., 2006; Lifschitz et al., 2006; Corbesier et al., 2007; Jaeger and Wigge, 2007; Mathieu et al., 2007; Tamaki et al., 2007; Komiya et al., 2009; Yant et al., 2010).

Vernalization, a prolonged exposure to cold (corresponding to winter), is required by winter *Arabidopsis* accessions in order to flower. Through genetic analysis on certain natural ecotypes it has been shown that the vernalization requirement results from synergistic interaction between dominant alleles of two different loci *FRIGIDA* (*FRI*) and

Figure 1.3 Major flowering time pathways in *Arabidopsis.* A simple schematic representing four major genetic pathways regulating flowering time in *Arabidopsis.* Photoperiod and vernalization pathways mediate environmental responses to regulate flowering. Autonomous and gibberellin pathways function independently of environmental cues. Autonomous pathway promotes flowering in all conditions, and the gibberellin pathway is required for flowering in non-inductive short-day conditions.

Figure adapted and modified from (Möller-Steinbach et al., 2010).


Vernalization Pathway
Photoperiod Pathway
Autonomous Pathway
Gibberellins Pathway
Flowering repressors
Flowering activators
Floral integrators
Vegetative to reproductive phase transition

FLOWERING LOCUS C (*FLC*) (Napp-Zinn, 1957; Burn et al., 1993; Clarke and Dean, 1994; Koornneef et al., 1998; Lee et al., 1994; Sanda and Amasino, 1996; Johanson et al., 2000). *FRI* promotes the accumulation of *FLC* mRNA and represses the floral transition even in normally favorable conditions (Johanson et al., 2000). Vernalization decreases the accumulation of the *FLC* mRNA, and the decreased levels of *FLC* mRNA positively correlate with the transition to flower (Michaels and Amasino, 1999). FLC is a MADS domain transcription factor and holds a central role in the flowering pathway in *Arabidopsis* (Michaels, 1999; Sheldon, 1999). The expression of the flowering pathway integrators *FT, FD* and MADS domain protein encoding gene *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1 (SOC1)* are supressed by *FLC* and *SHORT VEGETATIVE PHASE (SVP),* another MADS domain containing transcription factor (Lee et al., 2007; Li et al., 2008).

In the absence of vernalization, the *FLC* expression is suppressed by a group of genes (*FCA, FY, FLD, FPA, FVE, LD*, and *FLK*) that have been genetically grouped under the autonomous pathway (Mouradov et al., 2002; Simpson and Dean, 2002). Recent molecular evidence suggests that the genes in the autonomous pathway function in concert instead of forming a typical linear pathway to repress *FLC* expression and to promote flowering (Möller-Steinbach et al., 2010). In addition, flowering is promoted in response to endogenous signals, such as the phytohormone gibberellin. Gibberellins are plant growth regulators which induce flowering by acting directly at the level of meristem identity genes (Blazquez et al., 1998; Blázquez and Weigel, 2000).

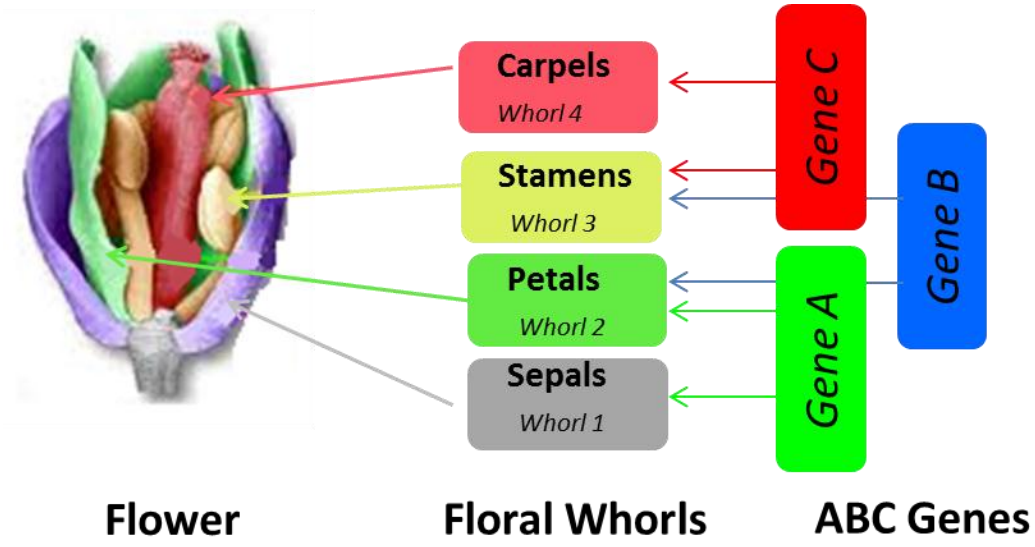
1.4 Development of flowers

Flowers have a simple stereotypical architecture despite the wide diversity of their shapes and forms. They are the reproductive part of the plant, consisting of a series of sterile organs surrounding a reproductive structure. The SAM converts to a reproductively determined inflorescence meristem after receiving the appropriate signals upon entry into the reproductive phase (Amasino, 2010). In *Arabidopsis*, the inflorescence meristem produces floral meristems that sequentially produce floral organs, the sepals, petals, stamens and carpels. These structures arise in concentric rings or whorls (Steeves and Sussex, 1989; Smyth et al., 1990). In *Arabidopsis* the first or the outermost whorl produces four sepals, which are leaf-like organs enclosing the developing flower bud. The second whorl produces four white petals alternating in position with the sepals. The third whorl produces six stamens, the male reproductive organ, which consists of a filament having an anther at the tip that produces pollen. The fourth or the central whorl produces the gynoecium, the female reproductive organ, which is formed by the fusion of two carpels. After fertilization, the gynoecium carrying the ovules produces seeds.

In 1991, Enrico S. Coen and Elliot M. Meyerowitz proposed the ABC model (Figure 1.4) to explain specification of whorls in developing flowers of *Arabidopsis* and snapdragons (*Antirrhinum majus*). The genes that control each whorl in *Arabidopsis* are referred to as homeotic genes. In the floral meristem, the 'A' genes, *APETALA1* (*AP1*) and *APETALA2* (*AP2*), are the first to become activated and they regulate the identity of the sepals, which are formed in the first whorl. In the second whorl, the 'A' and the 'B' genes *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) in combination regulate the identity of petals. Subsequently,

Figure 1.4. ABC model of flowering. The figure shows an *Arabidopsis* flower. The horizontal rectangular boxes indicating the whorls are color coded to match the floral organ they represent. The vertical rectangular boxes represent the A, B and C genes which act in combination to produce the flower.

(Figure modified from http://www.adonline.id.au/flowers/floral-identity/)


Carpels
Whorl 4
Stamens
Whorl 3
Petals
Whorl 2
Sepals
Whorl 1
Gene C
Gene A
Gene B
Flower
Floral Whorls
ABC Genes

the ‘B’ and the ‘C’ genes (*AGAMOUS, AG*) in combination, regulate the identity of stamens in the third whorl. ‘C’ gene controls the identity of carpels in the fourth whorl (Coen and Meyerowitz, 1991; Weigel and Meyerowitz, 1994). *AG* is the only characterized C-class gene in *Arabidopsis.* It is activated at a very early stage of development by genes *LEAFY* (*LFY*) and *WUSCHEL* (*WUS*) (Busch et al., 1999; Lenhard et al., 2001; Lohmann et al., 2001). Even though *AG* is expressed at a very early stage of floral development, little is known about the genes that function either together with or downstream of it. This could be because of redundancy or lethality in the *AG* pathway (Jack, 2002). Mutant screens to identify C organ identity genes in a wild-type background always resulted in mutants with either defects in floral determinacy or stamen-to-petal and carpel-to-sepal were all *ag* alleles. In a genetic screen aimed to isolate extragenic mutations that enhanced the weak *ag* allele, *ag-4*, *HUA1* and *HUA2* were identified as genes that participate in the specification of stamen and carpel identities and also in the control of floral determinacy (Chen and Meyerowitz, 1999). *HUA1* codes for an RNA-binding protein with six tandem CCH zinc fingers (Li et al., 2001) and *HUA2* codes for a novel protein. It has been shown that *HUA2* affects *AG* pre-mRNA processing in certain genetic backgrounds (Cheng et al., 2003; Wang and Grbic, unpublished).

1.5 Pre mRNA-processing

Proper gene expression requires many steps, including chromatin remodeling, transcription, RNA processing and transport, translation and post translation modification. All these steps are interconnected both directly and indirectly to a certain degree (Orphanides and Reinberg, 2002). During transcription, genetic information is transferred from DNA to messenger RNA. Most eukaryotic genes consist of blocks of

coding sequences (exons) separated from each other by blocks of non-coding sequences (introns). The exons and the introns are arranged in an alternating pattern. When a gene containing introns is transcribed, the primary transcript (also referred to as pre-mRNA) contains both exonic and intronic sequences. The primary transcript undergoes four main changes before it is translated: (1) A 7-Methyl guanosine cap is added to the 5′ end of the primary transcript; (2) The primary transcript is cleaved towards the 3′ end; (3) A Poly (A) tail is added to the 3′ end of the transcript; and (4) The non-coding introns are spliced out of the transcript. The process of intron removal is called splicing. This process converts the pre-mRNA into a mature mRNA. Splicing of nucleosomal pre-mRNAs requires a large complex of proteins and RNAs called the spliceosome, which is made up of snRNPs (small nuclear ribonuclear proteins). SR (Serine-Argenine rich) proteins and non-snRNP protein factors such as hnRNP (Heterogeneous ribonucleoprotein particles) help the spliceosome enhance its splicing accuracy. SR proteins are essential in several steps of both constitutive and alternative splicing. The hnRNPs are complexes of RNA and proteins present in the cell nucleus during gene transcription and subsequent post-transcriptional modification of the newly synthesized RNA (pre-mRNA).

1.5.1 Pre-mRNA splicing in plants

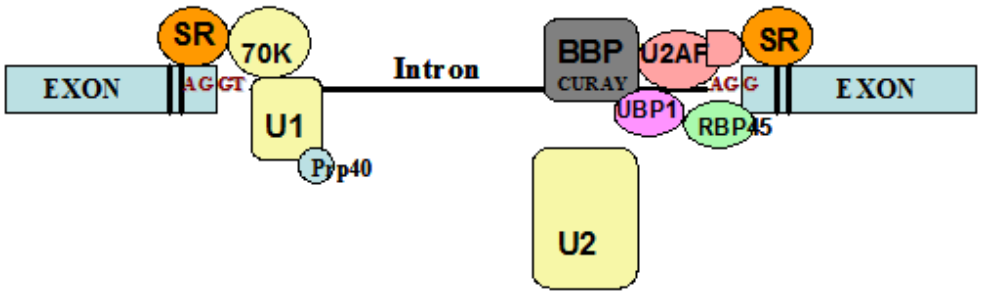
The expression of all intron-containing genes is affected by splicing. Approximately 20-30% of genes in *Arabidopsis* are alternatively spliced, suggesting the critical roles of splicing in gene expression (Campbell et al., 2006; Wang and Brendel, 2006). Many proteins are involved in the regulation of splicing. SR proteins assist and promote splicing and are referred to as splicing factors. Proteins that can either modify splicing factors or have a negative effect on splicing are referred to as splicing regulators. hnRNP

is one type of splicing regulator which can bind to pre-mRNA and block the binding of splicing factors (Wang and Brendel, 2004). Homologs of proteins and RNA factors involved in splicing in other eukaryotic systems have been identified in the *Arabidopsis* genome, suggesting that there is conservation of the basic steps in the splicing mechanism (McCullough and Schuler, 1997; Brown and Simpson, 1998; Lorković et al., 2000; Reddy, 2001). The conserved sequences found at 5′ and 3′ splice sites and branch points are also similar in plants, yeast and mammals. However, even with this high sequence similarity, it has been observed that plant introns are poorly spliced in mammals and *vice versa* (Belostotsky and Rose, 2005). One of the reasons for this incompatibility might be because of the complex splicing machinery in plants compared to yeast and mammals. *Arabidopsis* contains almost twice as many SR proteins compared to what is present in the human genome (Reddy, 2004). SR proteins are involved in multiple protein-protein interactions among themselves and other proteins and numerous constitutive and alternative splicing events (Graveley, 2000). Plant SR proteins are involved in recognition of the intron-exon boundary by binding to exonic and intronic splicing enhancers and silencer signals. They also help in recruitment of spliceosomal proteins onto pre-mRNA. Unlike vertebrates, plant exons are relatively short and are usually separated by large introns and are identified either by intron-binding factors or through a combination of exon and intron definition (McCullough and Schuler, 1997; Brown and Simpson, 1998; Lorković et al., 2000; Reddy, 2001). The splice site consensus sequences AG/GTAAGT and 3′ splice site consensus sequence TGCAG/G of plants are similar to that of vertebrates. Plant introns are strongly biased towards UA or U rich sequences and the 3′ end of the introns are rich in uridine residues, particularly in

dicots. These U rich sequences play an important role in the recognition of plant introns (Goodall and Filipowicz, 1989; Brown and Simpson, 1998).

Most of the research on pre-mRNA processing in plants has been focused on the role of UA rich sequences in the intron (Figure 1.5). Mutation analysis has shown that uridines are more essential for splice site selection than adenines (Merritt et al., 1997; Ko et al., 1998). Several hnRNPs are associated with recognition of the U rich sequences that facilitate spliceosomal association. UBP1, oligouridylate-binding nuclear three-RBD protein, was the first plant hnRNP to be characterized. *In vitro* studies have revealed that UBP1 has binding specificity for oligouridylates and that it interacts with U rich intron and 3′-UTR sequences. Overexpression of UBP1 in protoplasts increases the splicing of suboptimal introns and increases steady-state levels of transcripts (Lambermon et al., 2000). Several additional hnRNPs have been identified, UBP1 associate protein 1 (UBA1) , UBP1 associate protein 2 (UBA2), RNA BINDING PROTEIN 45 (RBP45) and RNA BINDING PROTEIN 47 (RBP47) (Lambermon et al., 2000; Lorković et al., 2000; Lambermon et al., 2002). Experiments performed on UBP1, UBA1 and UBA2 suggest that they are components of a complex which regulates the turnover of mRNAs in the nucleus (Lambermon et al., 2002). RBP45 and RBP47 also show specificity for oligouridylates and are associated with the poly(A)$^+$ RNA. But unlike UBP1, RBP45 and RBP47 do not enhance mRNA splicing and accumulation when transiently overexpressed in protoplasts (Lorković et al., 2000).

Figure 1.5. Schematic showing cis elements and trans-activation factors involved in the splicing of plant introns. Exons are shown in blue boxes and the intron is shown on the black line. 5' and 3' consensus nucleotides at the splice junction are shown along with branch point sequence CURAY. Binding positions of U1 and U2 SnRNPs in the intron are shown. Figure adapted and modified from Lorković et al. (2000).


SR
70K
EXON
AG GT
U1
Pyp40
Intron
BBP
CURAY
U2AF
UBP1
RBP45
SR
AG G
EXON
U2

1.6 C-terminal repeat domain of RNA polymerase II

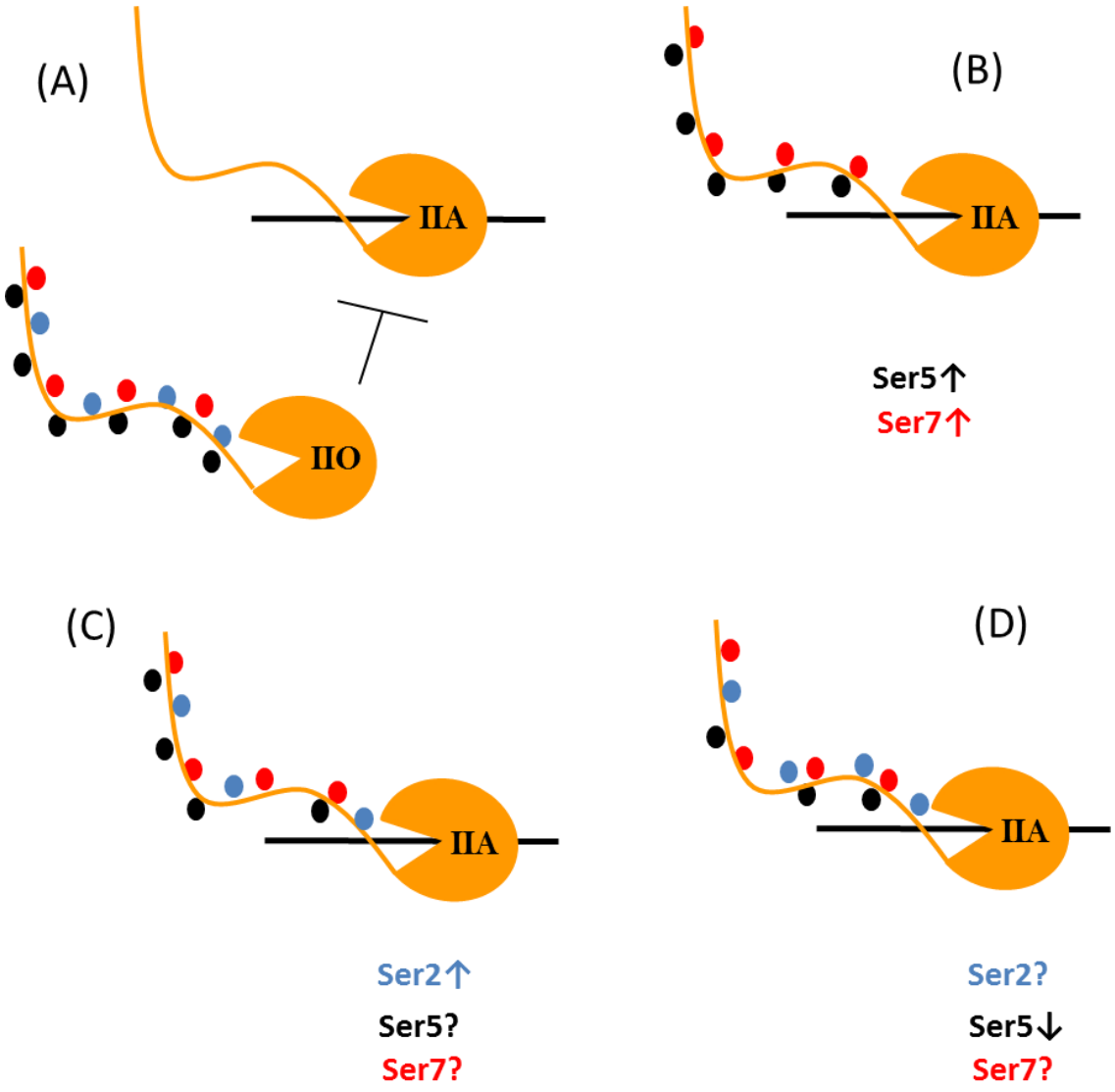
Eukaryotic RNA polymerase II (RNAP II) is responsible for synthesizing all protein-coding genes and most non-protein RNAs like small nuclear RNAs (snRNA), small nucleolar RNA (snoRNA) and microRNA (miRNA). RNAP II plays a key role and is involved in regulatory control at all stages of the transcription cycle including initiation, elongation and termination via its unique C-terminal repeat domain (CTD). CTD is attached to and is an extension of the C terminus of the largest subunit of RNA polymerase II. The CTD contains tandem hepta-peptide repeats with a consensus sequence of Tyr1Ser2Pro3Thr4Ser5Pro6Ser7 that has been observed to undergo extensive post-transcriptional modification in a coordinate and complicated manner. One such modification is phosphorylation. The phosphorylation state of CTD is subjected to a continuous flux throughout the entire transcription cycle with multiple kinases and phosphatases working in a concert to maintain specific phosphorylation patterns on particular subsets of heptads. The reversible phosphorylation of proteins, especially of the heptads within the CTD, has been associated with the regulation of eukaryotic transcription that is accompanied with pre-mRNA processing. This reversible phosphorylation at positions Ser2, Ser5 and Ser7 of the heptads in CTD plays a key role in recruiting and binding specificity of a specific set of factors at specific stages of the transcription cycle (Hirose and Manley, 2000; Chapman et al., 2007; Egloff et al., 2007). Its list of binding partners, modifications, and associated processes has grown rapidly over the years (Bartkowiak et al., 2011). The number of heptad repeats varies from organism to organism and it appears to be linked with genome complexity; for example, yeast contains 26 repeats, there are 44 in *Drosophila* and 52 in humans (Allison et al.,

1988; Corden, 1990). The *Arabidopsis* CTD has 34 heptad repeats that are involved in recruiting and tethering the RNA processing factors to the transcription complex (Morris and Greenleaf, 2000; Carty and Greenleaf, 2002).The RNAPII is recruited at the site of the promoter when the CTD is unphosphorylated. Phosphorylation of CTD prior to the establishment of the pre-initiation complex (PIC), a large complex of proteins that is necessary for the transcription of protein-coding genes in eukaryotes, has an inhibitory effect on transcription (Hengartner et al., 1998). After the formation of PIC the CTD is phosphorylated at the Ser5 and Ser7 positions then elongation of pre-mRNA begins simultaneously with an increase in phosphorylation in Ser2, producing a hyperphosporylated form of the CTD (Komarnitsky et al., 2000). As the polymerase glides close to the 3′ end of the coding gene, Ser5-specific phosphatase decreases the Ser5 phosphorylation levels, while the Ser2 and Ser7 phosphorylation levels are left largely unchanged (Tietjen et al., 2010) (Figure 1.6).In terms of RNA processing events, the function of CTD has been associated with mRNA capping, splicing, 3′ end processing and termination. But its role has not been restricted to the above functions; it is associated with non-RNA processing associated functions like transcription activation, cotranscriptional chromatin modification, chromatin remodeling and genome stability which are all interlinked in regulating gene expression (Bartkowiak et al., 2011).

1.7 Chromatin modifications and its role in regulation of FLC

As mentioned in the previous sections, one of the major developmental transitions in the plant life cycle is transition to flowering. In *Arabidopsis*, chromatin modification plays a

Figure 1.6. Phospho-CTD cycle for protein-coding genes (A) an unphosphorylated CTD (IIA form) of RNAPII is recruited to the promoter (black line), a phosphorylation of CTD (IIO form) prior to preintiation complex has an inhibitory effect on transcription. (B) Soon after the preinitiation complex is formed, Ser5 (black) and Ser7 (red) of CTD are phosphorylated. (C) An increase in Ser2 phosphorylation (blue) produces the hyperphosphorylated form of the CTD, during elongation. (D) Towards the 3′ end of the gene, a decrease in Ser5 phosphorylation is observed. Figure adapted and modified from Bartkowiak et al. (2011).


(A)
IIA
IIO
(B)
IIA
Ser5↑
Ser7↑
(C)
IIA
Ser2↑
Ser5?
Ser7?
(D)
IIA
Ser2?
Ser5↓
Ser7?

key role in regulating flowering time through the controlled expression of key flowering regulatory genes, especially of *FLC*. The level of *FLC* is controlled by several conserved chromatin modifiers, plant-specific factors and long noncoding RNAs (lncRNAs). Studies performed on understanding the regulation of *FLC* have set-up a foundation for understanding chromatin-based control of other developmental genes. Several developmental genes in plants are regulated by chromatin modifications, such as nucleosome remodelling, DNA methylation, and histone modifications (Henikoff and Shilatifard, 2011; He, 2012). Certain modifications, like histone acetylation, histone H3 lysine-4 trimethylation (H3K4me3), H2B monoubiquitination (H2Bub1), and H3 lysine-36 di- and trimethylation (H3K36me2/me3) are associated with active gene expression, whereas histone deacetylation, H3 lysine-9 methylation, H3lysine-27 trimethylation (H3K27me3), and H2A monoubiquitination (H2Aub1) are linked with gene repression (He, 2009).

The role of *FRI* in chromatin-mediated regulation of *FLC* has recently been identified. Genetic screens for mutants that suppress *FLC* activation in *FRI*-containing lines have identified several genes. Loss-of function mutations in these genes lead to decreased levels of *FLC* and inducing the *FRI*-containing line to flower early (Amasino, 2010; Crevillén and Dean, 2011). *FRI* functions in association with RNA POLYMERASE II-ASSOCIATED FACTOR 1 COMPLEX (PAF1c), which is highly conserved from yeast to plants and humans, and associates with RNAPII during transcription (He et al., 2004; Yu and Michaels, 2010). PAF1c acts as a platform for the attachment of histone modifying enzymes during transcription and elongation.

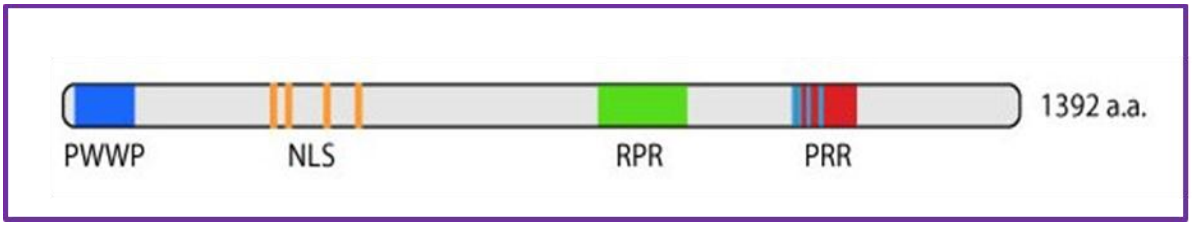
In *Arabidopsis*, PAF1c is composed of six subunits. Disruption of these units brings about reduction in H3K4me3, H3K36me2, and H3K36me3 on FLC chromatin and inhibits *FLC* expression in the *FRI* containing lines (He et al., 2004; Xu et al., 2008). FRI together with two plant specific factors FRL1 and FES1 and along with SUF4 and FLX (Michaels et al., 2004; Andersson et al., 2008) forms a putative transcription activator complex (FRIc)(reviewed in He, 2012). It has been identified that SUF4 within the complex is responsible for recognition of a *cis*-element in the *FLC* proximal promoter (Choi et al., 2011). Loss of function mutations in any of the subunits of the FRIc complex leads to decreases in *FLC* expression resulting in an early flowering phenotype, suggesting that this complex is a *FLC* specific activator (reviewd in He, 2012). FRIc has also been associated with chromatin modifiers like EFS and SWR1c (Ko et al., 2010a; Choi et al., 2011). *FLC* is actively expressed after the modification of H3K4me3 on its locus by COMPASS-like H3K4 methyltransferase complexes. COMPASS is made up of four conserved subunits; a SET-domain H3K4 methyltransferase along with three structural components known as WDR5a, RBL and ASH2R in *Arabidopsis*. These subunits form a complex which provides a structural platform for H3K4 methylation. It has also been observed that a functional *FRI* is required for enrichment of WDR5a at *FLC* locus, suggesting that FRIc recruits or enriches multiple chromatin modifiers at *FLC* chromatin (Ko et al., 2010a; Choi et al., 2011). In the absence of an active *FRI* allele, *FLC* expression is repressed through chromatin silencing (He, 2012). *FLC* repression requires long noncoding RNAs (lncRNAs) that originate from an antisense *FLC* transcript (Liu et al., 2007). The 3'-end of the *FLC* antisense transcript is deferentially processed so that the utilization of the proximal poly(A) site, promoted by FPA, FCA, FY

and CstF, triggers localized histone demethylase activity and results in reduced *FLC* sense transcription (Liu et al., 2007; Liu et al., 2010). Along with chromatin modifiers and several other genes, *HUA2* has been identified as one of the genes required for both *FRI*-dependent and *FRI*-independent *FLC* expression (Doyle et al., 2005; Wang et al., 2007). The molecular mechanism in which *HUA2* regulates *FLC* expression is yet to be resolved.

1.8 Identification of *HUA2* in our laboratory

As mentioned in section 1.1.3, *HUA2* was first identified in a genetic screen aimed to isolate extragenic mutations that enhance the weak *ag* allele, *ag-4*. The two genes, *HUA1* and *HUA2*, were identified to participate in the specification of stamen and carpel identities and also in the control of floral determinacy (Chen and Meyerowitz, 1999). In our laboratory, *HUA2* was identified in *Arabidopsis* accession *Sy-0* and was initially referred to as *AERIAL ROSETTE 1* (*ART1*).Unlike most common laboratory strains of *Arabidopsis*, the morphology of the late flowering *Sy-0* accession is characterized by enlarged basal rosette leaves, formation of aerial rosettes in axils of cauline (stem) leaves and reversion of inflorescence and floral meristems (Poduska et al., 2003). Genetic analysis on *Sy-0* indicated that *Sy-0* alleles at *FRI*, *FLC* and *ART1* are required for the morphology of *Sy-0*. At the time of its identification, *ART1* was believed to be a new gene in the flowering pathway, and was identified to repress flowering by increasing the expression of *FLC* (Poduska et al., 2003). Later, cloning and sequencing of *ART1* suggested that it was an allele of *HUA2* (Wang et al., 2007)

Figure 1.7. Schematic of HUA2 protein with various conserved domains. Towards its N-terminus end it has a 54- amino acid PWWP domain (amino acids 30-84). Between amino acids 200-450 it has four NLS. The RPR domain is between amino acids 778-909. Towards the C-terminus domain of the protein there are regions rich in prolines (amino acids 1056-1280).


1392 a.a.
PWWP
NLS
RPR
PRR

1.9 Structure of HUA2 protein

HUA2 encodes a novel plant protein with some conserved domains. The sequence of HUA2 protein does not match any known proteins whose function has been characterized (Chen and Meyerowitz, 1999). At the amino-terminus end it has PWWP domain followed by four nuclear localization signals (NLS). It also contains a RPR domain followed by a proline rich region with 4 PPLP repeats and ends with amino acids rich in proline/serine towards the carboxyl-terminus end (Figure 1.7).

1.9.1 PWWP domain

The PWWP domain was first identified in WHSC1, a gene mapping to the Wolf-Hirschhorn syndrome critical region in humans. It is named after the central core motif Pro-Trp-Trp-Pro (PWWP) of this domain in WHSC1 (Ceulemans et al., 2000). The domain is made up of 100-130 amino acids and is often present in proteins which are known to associate with chromatin. It belongs to the Tudor domain 'Royal Family' which includes Tudor, chromodomain and the MBT domain (Maurer-Stroh et al., 2003). There are around 16 genes encoding PWWP-containing proteins in the *Arabidopsis* genome (Alvarez-Venegas and Avramova, 2012). The functional significance of this domain has not been identified in any plant protein so far. The current knowledge of this domain comes from work on proteins having this domain in humans and yeast. The sequence and the structural alignments of PWWP domain indicate that this domain is structurally similar to the other 'Royal Family' members. Like the other members of the 'Royal Family' it contains an aromatic ring. Structural comparison of the architecture of the PWWP domain with the chromodomain, MBT and Tudor domains shows that PWWP, MBT and Tudor all have a 5-β-strand canonical core, while the chromodomain consists

of three β-strands and one α-helix (Wu et al., 2011). The yeast protein Pdp1 has a PWWP domain that has been shown to bind monomethylated histone H4K20. Pdp1 interacts with yeast H4K20 methyltransferase SET9, which is recruited to the H4K20me1 chromatin region through PWWP domain of Pdp1. This leads to an increase in the concentration of SET9 on chromatin which carries out trimethylation of histone H4 (H4K20) (Wang et al., 2009). Other human proteins with a PWWP domain, like BPRF1 and DNMT3A, have been shown to bind H3K36me3 through their PWWP domains (Dhayalan et al., 2010; Vezzoli et al., 2010). The PWWP domains in BRPF1, BRPF2, HDGF2, MUM1 and the N-terminal PWWP domains of WHSC1 and WHSC1L1 have been shown to have weak binding affinity to histones with H3K36, K3K79 or H4K20 methylation (Wu et al., 2011). Studies done so far on different proteins with PWWP domain suggest that PWWP domain shows methyl lysine histone binding ability. Unlike other PWWP domains which are around 100–130 amino acids long, the PWWP domain in HUA2, which is present at the N terminus end, is only 54 amino acids in length. It shows significant similarity to parts of proteins having PWWP domains.

1.9.2 Nuclear localization signals (NLS)

Transport of nuclear proteins is triggered by the presence of a nuclear localization signal (NLS), a short amino acid sequence contained within nuclear protein domains (Cokol et al., 2000). Presence or absence of an NLS is a key factor determining if nuclear transport will occur. If an NLS is deleted, transport will not occur and if an NLS sequence is added to a non-nuclear protein, transport will occur. This property has been used to help elucidate NLS motifs in nuclear proteins (Tinland et al., 1992; Moede et al., 1999). The exact number of NLS and their variety is not well understood and it is uncertain if a

consensus NLS motif exists. Many NLS motifs contain a large number of positively charged residues that are essential for nuclear transport as mutations of these positive residues prevent entry of the protein into the nucleus. Positively charged residues facilitate nuclear transport by binding to importins (Conti et al., 1998). However, NLS motifs have been identified that contain large amounts of glycine and a low number of positively charged residues as well (Bonifaci et al., 1997). The monopartite and bipartite NLS motifs are among the best experimentally characterized glycine-rich NLS (Boulikas, 1993). The monopartite motif usually contains a helix-breaking residue followed by a series of basic residues. The bipartite motif contains two series of basic residues separated by 9–12 residues but lacks a helix-breaking residue. It is questioned if these NLS represent a universal NLS structure because these simple characteristics are observed in many non-nuclear proteins. Also, many nuclear proteins do not follow the above consensus rules (Hsieh et al., 1998; Truant and Cullen, 1999; Irie et al., 2000). Therefore, although knowledge has been gained on the basic properties and characteristics of NLS, information on the universal patterns and structures of NLS is needed.

1.9.3 RPR domain

RPR domain is also referred to as CTD-Interaction Domain (CID) and is a common domain involved in phosphopeptide binding and regulation of RNA metabolism (Patturajan et al., 1998; Doerks, 2002). It has been shown to be involved in interaction with the CTD of RNA polymerase II and is present in several proteins involved in regulation of nuclear pre-mRNA (Doerks, 2002). The structure and function of this domain has been intensively studied in Pcf11 protein of *Schizosaccharomyces pombe*.

Pcf11 is one of the subunits of the yeast Cleavage Factor IA (CFIA) which is required for 3' end processing. Pcf11 has been shown to preferentially bind to the phosphorylated CTD (Rodriguez et al., 2000; Barillà et al., 2001; Dichtl et al., 2002). The crystal structure of the CID from Pcf11 suggests that this motif forms a compact domain of eight α-helices arranged in three antiparallel pairs. NMR data from Noble et al. (2005) suggest that the CTD binding proteins scan along the CTD from one repeat to the next until the assembly of the complete CFIA complex with the appropriate combination of post-transcriptional modifications stabilizes the binding.

1.9.4 The carboxyl-terminus proline rich region

The exact function of proline-rich regions in proteins is still not clear but it has been suggested that in many proteins it appears that they mediate functionally important binding interactions (Williamson, 1994). In 1997, Bedford et al. showed that the proline rich region having PPLP repeats is responsible for interactions with proteins containing a domain with two highly conserved tryptophans, the WW domain. The amino acid proline has been critical among the primary structures of many ligands for protein–protein interactions. Ligand sequences that are rich in prolines are preferred by SH3, WW, and several other protein interaction domains (Kay et al., 2000).

1.10 Role of *HUA2* in *AG* and *FCA* pre-mRNA processing

Work from our laboratory on *HUA2* suggests that it is required for the efficient splicing of *AG* (Wang and Grbic, unpublished results). *AG* produces multiple splice variants that include a short transcript that terminates in the second intron due to the selection of a proximal nondescript signal. The accumulation of this short transcript was higher in

hua2-7 null background when compared with wild-type. The same pattern was also observed in a partial loss of function *hua2-1* mutant suggesting that *HUA2* is required for the efficient splicing and/or 3' end processing of *AG* transcript (Sajja, Wang and Grbic, unpublished results). Like *AG*, *FCA*, which contains 20 introns, is alternatively spliced. *FCA* produces four different transcripts, designated α, β, δ, and γ, which account for <1, 55, 10, and 35% of splice-variant transcript abundance, respectively (Macknight et al., 1997). The full-length protein, which includes two RNA binding motifs and a WW domain, that is known to participate in protein-protein interaction, is encoded by the γ transcript. The most abundant transcript, β, is produced from cleavage and polyadenylation within the third intron, which produces a truncated transcript that encodes a protein lacking RNA binding and WW domain (Macknight et al., 1997). In *fca-1* and *fy-1* the levels of γ transcript are greater than β transcript (Quesada et al., 2003). Our laboratory demonstrated that in *hua2-1* mutant lines accumulation of β transcript increases further, suggesting that *HUA2* is required for proper *FCA* processing and accumulation of γ transcript (Sajja and Grbic, unpublished results).

1.11 HUA2 is required for H3K4 trimethylation at FLC locus

It was previously suggested that *HUA2* is required for the increased levels of *FLC* transcript (Doyle et al., 2005; Wang et al., 2007). Regulation of gene expression occurs at initiation and elongation steps during transcription. After transcribing a few nucleotides, RNA Pol II pauses until a histone 3 Lys4 (H3K4) is methylated by a methyl transferase enzyme. Components of the *Arabidopsis* Paf1 complex which includes *ELF7* is required for this methylation event (He et al., 2004). In late flowering line *FRI* Col, a high level of trimethylation was observed at the H3-K4 position. However, our laboratory

demonstrated that this increase in trimethylation levels was lost in *FRI hua2-3* background, indicating that a functional *HUA2* is required for the trimethylation of histone H3 at *FLC* locus. This suggests that *FRI* mediated increase in H3K4 methylation is dependent on *HUA2* along with components of Paf1 complex like *ELF7* (Grbic, unpublished).

1.12 Identification of HULKs

The existence of three additional genes with high sequence similarity to *HUA2* was identified when a BLAST search was conducted against the TAIR database with both the nucleotide and protein sequences of *HUA2*. These genes were named *HULK (HUA2 LIKE)* genes and together with *HUA2* the family is referred to as the *HULK* gene family. *HUA2* and *HULK1* genes are located on chromosome 5, *HULK2* on chromosome 2 and *HULK3* on chromosome 3. The amino acid sequences of *HULKs* share high similarity with HUA2 and the protein architecture and conserved domain structures of the HULKs are the same as seen in HUA2 (Figure 1.8). All four gene family members show a similar expression pattern, which is concentrated at the vegetative and reproductive apices (Challa, unpublished). Figure 1.9 shows a phylogenetic tree created with the PWWP-containing proteins from phylogenetically disparate plant species. The tree is divided into two main branches; one branch contains sequences homologous to HUA2 and HULK1 protein and the other branch contains sequences homologous to HULK2 and HULK3. The *HULK* genes seem to have separated in all the common ancestors of Angiosperms. The shared expression patterns among the *HULK* gene family at tissue level along with similar

Figure 1.8. Comparison of functional domains within HULK protein family.

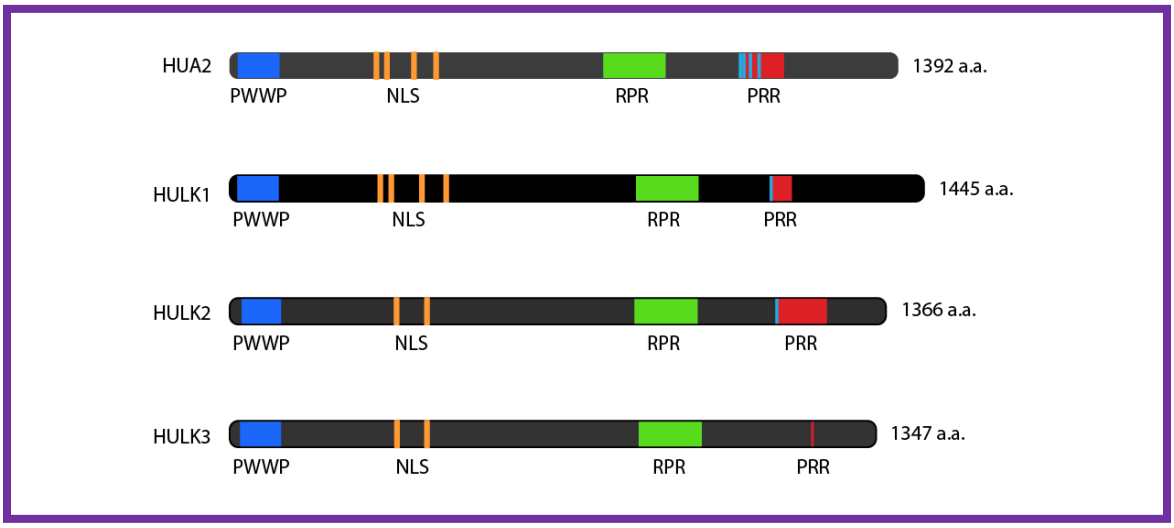
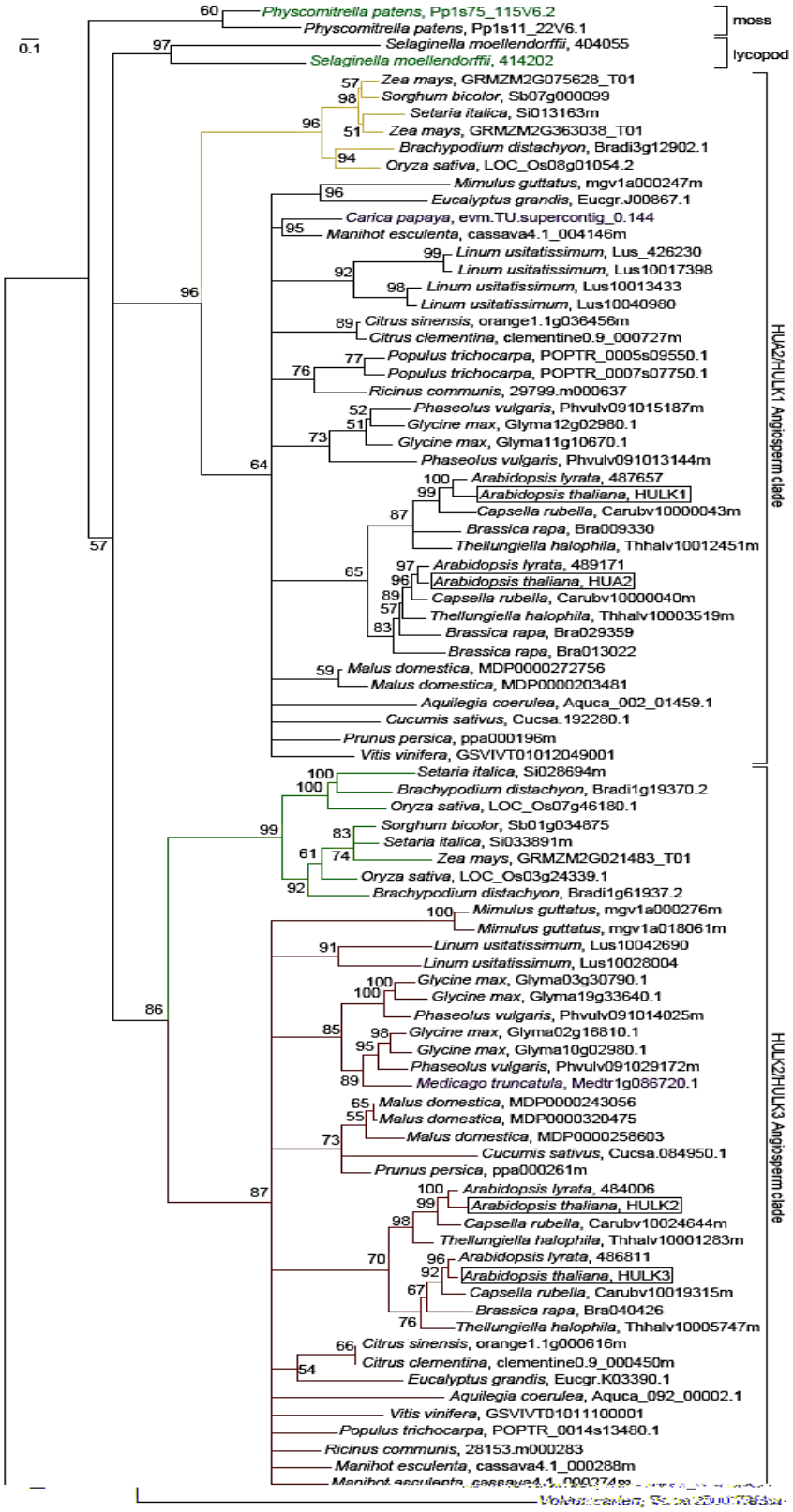

HUA2
1392 a.a.
PWWP
NLS
RPR
PRR
HULK1
1445 a.a.
PWWP
NLS
RPR
PRR
HULK2
1366 a.a.
PWWP
NLS
RPR
PRR
HULK3
1347 a.a.
PWWP
NLS
RPR
PRR

Figure 1.9. Rooted phylogram of amino acid sequences of PWWP-containing proteins of 29 Embryophyta species and *V. carteri* (as an outgroup) based on maximum likelihood. Sequences highlighted in green and an outgroup sequence were manually added to the data set. Sequences highlighted in purple are single representatives of a given species. Sub-trees are color-coded as the following: yellow – HUA2/HULK1 homologues in grasses, blue – HUA2/HULK1 homologues in dicots, green – HULK2/HULK3 homologues in grasses, red – HULK2/HULK3 homologues in dicots. Scale bar is a number of amino acid substitutions per site. Support values are puzzle support values.


0.1
Physcomitrella patens, Pp1s75_115V6.2
Physcomitrella patens, Pp1s11_22V6.1
Selaginella moellendorffii, 404055
Selaginella moellendorffii, 414202
moss
lycopod
Zea mays, GRMZM2G075628_T01
Sorghum bicolor, Sb07g000099
Setaria italica, Si013163m
Zea mays, GRMZM2G363038_T01
Brachypodium distachyon, Bradi3g12902.1
Oryza sativa, LOC_Os08g01054.2
Mimulus guttatus, mgv1a000247m
Eucalyptus grandis, Eucgr.J00867.1
Carica papaya, evm.TU.supercontig_0.144
Manihot esculenta, cassava4.1_004146m
Linum usitatissimum, Lus_426230
Linum usitatissimum, Lus10017398
Linum usitatissimum, Lus10013433
Linum usitatissimum, Lus10040980
Citrus sinensis, orange1.1g036456m
Citrus clementina, clementine0.9_000727m
Populus trichocarpa, POPTR_0005s09550.1
Populus trichocarpa, POPTR_0007s07750.1
Ricinus communis, 29799.m000637
Phaseolus vulgaris, Phvulv091015187m
Glycine max, Glyma12g02980.1
Glycine max, Glyma11g10670.1
Phaseolus vulgaris, Phvulv091013144m
Arabidopsis lyrata, 487657
Arabidopsis thaliana, HULK1
Capsella rubella, Carubv10000043m
Brassica rapa, Bra009330
Thellungiella halophila, Thhalv10012451m
Arabidopsis lyrata, 489171
Arabidopsis thaliana, HUA2
Capsella rubella, Carubv10000040m
Thellungiella halophila, Thhalv10003519m
Brassica rapa, Bra029359
Brassica rapa, Bra013022
Malus domestica, MDP0000272756
Malus domestica, MDP0000203481
Aquilegia coerulea, Aquca_002_01459.1
Cucumis sativus, Cucsa.192280.1
Prunus persica, ppa000196m
Vitis vinifera, GSVIVT01012049001
HUA2/HULK1 Angiosperm clade
Setaria italica, Si028694m
Brachypodium distachyon, Bradi1g19370.2
Oryza sativa, LOC_Os07g46180.1
Sorghum bicolor, Sb01g034875
Setaria italica, Si033891m
Zea mays, GRMZM2G021483_T01
Oryza sativa, LOC_Os03g24339.1
Brachypodium distachyon, Bradi1g61937.2
Mimulus guttatus, mgv1a000276m
Mimulus guttatus, mgv1a018061m
Linum usitatissimum, Lus10042690
Linum usitatissimum, Lus10028004
Glycine max, Glyma03g30790.1
Glycine max, Glyma19g33640.1
Phaseolus vulgaris, Phvulv091014025m
Glycine max, Glyma02g16810.1
Glycine max, Glyma10g02980.1
Phaseolus vulgaris, Phvulv091029172m
Medicago truncatula, Medtr1g086720.1
Malus domestica, MDP0000243056
Malus domestica, MDP0000320475
Malus domestica, MDP0000258603
Cucumis sativus, Cucsa.084950.1
Prunus persica, ppa000261m
Arabidopsis lyrata, 484006
Arabidopsis thaliana, HULK2
Capsella rubella, Carubv10024644m
Thellungiella halophila, Thhalv10001283m
Arabidopsis lyrata, 486811
Arabidopsis thaliana, HULK3
Capsella rubella, Carubv10019315m
Brassica rapa, Bra040426
Thellungiella halophila, Thhalv10005747m
Citrus sinensis, orange1.1g000616m
Citrus clementina, clementine0.9_000450m
Eucalyptus grandis, Eucgr.K03390.1
Aquilegia coerulea, Aquca_092_00002.1
Vitis vinifera, GSVIVT01011100001
Populus trichocarpa, POPTR_0014s13480.1
Ricinus communis, 28153.m000283
Manihot esculenta, cassava4.1_000288m
HULK2/HULK3 Angiosperm clade

protein architecture suggests that they might function redundantly, whereas the separation of the HULKs within *Arabidopsis* and other plant taxa observed in the phylogenetic tree suggests functional diversification of the gene family (Challa, unpublished).

1.12.1 Phenotypic characterization of HULK mutants

To determine the extent of functional redundancy among the *HULK* gene family members, our laboratory created single, double and triple mutants lines of the HULK genes. Homozygous quadruple mutants were never rescued, suggesting that the gene family is essential for plant survival. The flowering time of all single mutants except *hua2-7* were the same when compared to control Col. *hua2-7* has been shown to flower slightly earlier than the wild type (Wang et al., 2007). However, when compared to the wild type, double mutants *hua2-7/hulk1* were early flowering, whereas *hulk2/hulk3* were late flowering, suggesting that the combination of double mutants from members of the same clade have an opposite effects on the flowering time. Cross-clade double mutants flowered at the same time as the control. In triple mutants, the combinations lacking *HUA2* were early flowering but the combinations with a functional *HUA2* were late flowering. Along with the variation observed in flowering time, compound mutants in the *HULK* gene family displayed morphological abnormalities. In double mutant combination, *hua2-7/hulk1* was the only one which showed a prominent phenotype when compared with controls. Early flower sterility with altered phylotaxy and embryo defects, which were never observed in single mutants and the wild type, were observed in *hua2-7/hulk1* mutant lines. The embryo defect was more pronounced in triple mutant *hua2-7/hulk1/hulk2*. Triple mutants *hua2-7/hulk1/hulk2* and *hua2-7/hulk1/hulk3* showed defects in seed germination, pollen viability and plant growth (Challa, unpublished).

While the double mutants show various morphological abnormalities, single mutants display normal development, suggesting that the *HULK* genes act redundantly. The pleiotropic phenotypes observed in the compound mutants and the inability to recover a quadruple mutant suggests that the *HULK* gene family members participate in an important cellular process required for plant development.

1.12.2 *HULKs* are associated with regulation of *FLC* expression

Like *HUA2,* the *HULK* genes affect the regulation of *FLC* expression. In the absence of *HUA2, hulk* mutants display early flowering, which is correlated with reduced *FLC* levels. When compared to wild type, low levels of *FLC* are observed in the double mutant *hua2-7/hulk1* (which is early flowering), whereas *hulk2/hulk3* (which is late flowering) shows high levels of *FLC* (Challa, unpublished). This effect further suggests that the gene pairs *HUA2/HULK1* and *HULK2/HULK3* have opposite roles in regulating *FLC* levels.

1.13 Hypothesis and Objective

Hypothesis:

Based on the protein structure of HUA2 protein family members and their effect on *AG*, *FCA* and *FLC* expression, I hypothesize that the members of HULK protein family localize in the nucleus where they associate with the regulation of gene expression by interacting with RNA binding proteins and chromatin modifiers.

Objectives:

The overall aim of this study is to understand the molecular mechanism through which the *HULK* gene family affects cellular function and development. Using different protein-protein interaction techniques such as the yeast two-hybrid assay and GST pull-down assay, my first objective is to identify proteins interacting with HUA2 followed by mapping the domain responsible for the protein interaction. The *HULK* genes encode proteins with domain architectures similar to HUA2, indicating the HULK proteins might potentially interact with the same proteins as HUA2. For the following reason, my second objective is to determine if the HULK proteins interact with same proteins as HUA2. All the members of the HUA2 protein family have putative nuclear signals; therefore, my third objective is to determine their cellular localization using transient expression assays. Mutation studies suggest that HUA2 is the main member of the family, hence I would like to concentrate in this study on understanding the independent domains of HUA2 using site directed mutagenesis and genetic complementation tests. I hope that my research will shed more light on understanding the molecular mechanism of HUA2 and its protein family and further help us understand their role in plant development.

Chapter Two: Materials and Methods

2.1 Plant material

Arabidopsis thaliana seeds were obtained from the *Arabidopsis* Biological Resource Centre (Ohio State University, Columbus, Ohio, USA). Double mutants used for domain characterization and complementation studies were created by Dr. Sathya Challa (Challa Ph.D. thesis, 2009).

2.2 Cultivation of plants in soil

Seeds of *Arabidopsis* were surface sterilized by soaking in 50% ethanol for one minute and then washing for 10 minutes in 50% bleach with 0.1% triton X-100 (Sigma). Seeds were then rinsed with sterile water five times to completely remove all bleach residues. To achieve uniform distribution, seeds were suspended in a 0.1% agar solution and stratified at 4 °C for 2–4 days to break dormancy and achieve synchronous seed germination.

The seeds were grown in PRO-MIX soil (Rivière-du-Loup, Québec, Canada). To control pests, the soil was pre-soaked with distilled water and then autoclaved for 20 minutes at 120 °C and 15 psi. Once autoclaved, the soil was moistened with enough water so that the soil was wet, but not dripping. Rectangular pots four inches wide which were able to accommodate 4–8 plants were used. The soil was packed loosely into pots and filled to approximately one and a half inch from the top. Eighteen square pots were placed in 55 cm x 28 cm plastic trays. Four to ten seeds were placed onto the soil in each pot using a Pasteur pipette or tweezers. Seedlings were thinned two days after germination to

prevent overcrowding. Plants were watered as needed and grown at 23°C, 75% humidity and under 16/8h light/dark regime with 100–150 μE/m^2/s.

2.2.1 Cultivation of plants on solid media

To select transformed plants, sterilized seeds were placed on sterile 1X MS-agar plates (Murashige and Skoog basal medium 0.433%, sucrose 2%, Agar 0.8%) having appropriate antibiotics. After depositing seeds on Petri dishes with agar, plates were sealed with Parafilm to prevent desiccation. The plates were then kept at 3–4 °C for 2–4 days to break dormancy. Plates were then moved to growth chambers and incubated at 23–25°C with a light intensity of 100–150 μE/m^2/s. Seedlings at 2-3 leaf stage were carefully transplanted into soil without damaging the roots. Plants were then grown under the growth chamber conditions stated in (Section 2.2).

2.3 DNA sequencing

The sequences of all genes/alleles in the constructs prepared for this project were verified by sequencing. DNA sequencing was performed at the Robarts Research Institute in London, Ontario, Canada and Max Planck Institute in Tubingen, Germany.

2.4 Bacterial and yeast strains

For general cloning procedures, *Escherichia coli* strains DH5α and DB3.1 (Invitrogen) were used. For protein expression purposes, *E. coli* strain BL21 (DE3) pLysS (Novagen) was used. For stable transformation of *Arabidopsis*, *Agrobacterium tumefaciens* strains ASE with kanamycin and chloramphenicol or GV3101 with rifampicin and gentamicin resistance (Koncz and Schell, 1986) were utilized. For yeast, strains MaV203

(Invitrogen) and PJ69-4A (kindly provided by Dr. Culbertson, University of Wisconsin, Madison, Wisconsin) were used.

2.4.1 Preparation of competent cells DH5 α, DB3.1 and BL23 (pLysS)

The procedure to make competent cells was adapted from the Promega *Protocols and Applications Guide*. Using a sterile inoculation loop the cells from frozen glycerol stock were streaked on Luria-Bertani (LB) agar plates (Bacto-tryptone 1%, yeast extract 0.5%, NaCl 1%, Agar 1.5%, pH 7.5) and incubated at 37°C for 12–16 hours. A single colony was picked and inoculated in 5 ml of LB medium incubated overnight at 37°C on a mechanical shaker at 220 rpm for starter culture. The following day 1% of the starter culture was diluted with 250 ml of LB medium in 1 L Erlenmeyer flask. The cells were allowed to grow until they reached an OD of 0.4–0.6 at $A_{600.}$ The cells were then cooled on ice for 5 minutes and pelleted by centrifugation at 4,500x g for 5 minutes at 4 °C in 250 ml centrifuge tubes in a large rotor (Sorvall GSA, Beckman JA-14). The cell pellet was then gently resuspended in 100 ml of ice-cold TFB1(RbCl 100mM, $MnCl_2$-$4H_2O$ 50 mM, Potassium Acetate 30 mM,$CaCl_2$-$2H_2O$ 10 mM, Glycerol 15% final, pH 7.5) The resuspended cells were transferred to smaller centrifuge tubes and left on ice for 5 minutes then centrifuged at 4,500x g for 5 minutes at 4 °C. The cell pellet was gently resuspended in 10 ml of ice-cold TFB2 (0.2M MOPS 10 mM, RbCl 10 mM,$CaCl_2$-$2H_2O$ 10mM, Glycerol 15%, pH 6.5) 100µL of the resuspended cells were dispensed into microfuge tubes, snap frozen in liquid nitrogen and stored at -80 °C.

2.4.2 Preparation of electro-competent *A. tumefaciens* cells

A. tumefaciens strain GV1301 was grown overnight at 28 ℃ in 5 ml of LB containing the appropriate antibiotics. Next, 500 µl of overnight culture was added to 500 ml LB liquid media. This was grown to an OD_{600} of approximately 0.6–0.8. The cells were washed by centrifuging three times with 10% ice-cold glycerol at 5000 rpm for 10 min at 4 ℃. The resulting pellet was resuspended in 2 ml of ice-cold 10% glycerol. 100µL of the resuspended cells were aliquoted into microfuge tubes, dropped in liquid nitrogen to quick-freeze and stored at -80 ℃ for future use.

2.4.3 Transformation of DH5α, DB3.1 and BL21 competent cells by heat shock

Transformation of DH5α, DB3.1 and BL21 competent cells by heat shock was performed as described in Sambrook et al. (1989). All transformations were performed using the heat shock procedure. Microfuge tubes containing competent cells were thawed on ice for 15 minutes after adding 2-5 µl of plasmid DNA. Next, the tubes were heat shocked at 42 ℃ on a heat block for 90 seconds and immediately placed back on ice for two minutes. Liquid LB was added (1 ml) to the tube and placed in a shaker at 37 ℃ for recovery. The tube was later spun at 7,000 rpm for 4 minutes and 800 µl of supernatant was discarded. The pellet was resuspended in the remaining supernatant and 50–100 µl of this mixture was plated on LB agar plates with appropriate antibiotics. The plates were incubated at 37 ℃ overnight.

2.4.4 Transformation of agrobacterium cells by electroporation

1-2 µl of plasmid DNA was added to the tube containing the electro-competent agrobacterium cells and mixed well. The mixture was transferred into sterile pre-chilled

disposable cuvettes (Eppendorf). Electroporation was carried out using the Eppendorf Multiporator® run on Eukaryotic mode with pulse voltage. Immediately after electroporation, 1 ml of LB media was directly added to each cuvette. Subsequently, cells were incubated in a shaker at 28 ºC for 1 h. Cells were then pelleted and plated on LB plates containing appropriate antibiotics.

2.4.5 Antibiotics in growth media

Antibiotic stock solutions used for this study are listed in table 2.1

2.4.6 Yeast competent cell preparation and transformation

Using a sterile inoculation loop, yeast cells (MaV203/ PJ69-4A) were streaked on YPAD agar plates ((Bacto-yeast extract 1%, Bacto-peptone 2%, Dextrose 2%, Adenine sulfate 0.01%, agar 2%, pH 6.0) and incubated for 24 hr at 30ºC. A single colony was picked from each plate and inoculated in 2XYPAD broth (2 ml per transformation) and incubated in a shaker at 30 ºC at 220 rpm. For each transformation, 2 ml of this culture was centrifuged at 13,000 rpm for 1 minute and the supernatant was discarded. To this pellet 240 µl of PEG 3500 50% w/v, 36 µl of 1.0 M LiAc, 50 µl of boiled SS-Carrier DNA (2 mg/ml) and 34 µl of plasmid DNA (0.1 to 1 µg) in sterile autoclaved double-distilled water was added. After adding all the reagents, the pellet was thoroughly resuspended and the cells were incubated in a water-bath at 42 ºC for 120 minutes. The cells were then pelleted by centrifuging at 13,000 rpm for 1 minute followed by removal of the supernatant. The pellet was resuspended in 200 µl of sterile distilled water and was plated on the appropriate Synthetic Complete Medium agar plate.

Table 2.1. List of Antibiotic stock solutions

Antibiotic	Stock
Ampicillin in water	100 mg/ml
Carbenicillin	100 mg/ml
Chloramphenicol in 95% ethanol	100 mg/ml
Gentamycin in water	50 mg/ml
Kanamycin in water	50 mg/ml
Rifamycin in 95% ethanol	25 mg/ml
Spectinomycin in water	80 mg/ml

2.5 Cloning

2.5.1 Cloning of genes and alleles for yeast two-hybrid assay

Table 2.2 contains list of genes/alleles cloned into yeast two-hybrid vectors used for the yeast-two hybrid assay. Entries printed in blue text indicate constructs I created. Entries printed in red were cloned by previous lab members or were obtained from an outside research groups. The allele *hua2-5* was amplified from HUA2pDBLeu (Sajja Ph.D. Thesis,2009) using primers hua2-5 Fwd and hua2-5 Rev having integrated *SalI* and *NotI* restriction sites, respectively (refer to Table 2.3 for all primer sequences used in the yeast two-hybrid assay). The cDNA was amplified with *Phusion*™ High-Fidelity DNA polymerase and cloned into *Sma I* site of pBLUESCRIPT cloning vector. The *hua2-5* alleles were later sub-cloned into pDB Leu using *SmaI* and *NotI* to create a hua2-5 pDBLeu construct which was translationally fused to GAL4 DNA binding domain (GAL4DB-hua2-5).

The WW domain containing part of *ArabidopsisPRP40* (PRP40(ww)) was synthesized using RNA extracted from *Arabidopsis* accession Columbia (Col-0) using a Qiagen RNA extraction kit following the manufacturer's instructions. RNA (1 µg) was used to prepare complementary DNA (cDNA) using SuperScript® II Reverse Transcriptase (*RT*) from Invitrogen following the manufacturer's instructions. PRP40 (ww) was amplified using primer pair PRP40 (ww) pDB Fwd/Rev (listed in Table 2.2). The amplified product was cloned into pGEM®-T Easy (Promega) vector following the manufacturer's instructions. Later *PRP40(ww)* was sub-cloned into the *SalI* and *NotI* sites of pDBLeu to create GAL4DB-PRP40(ww). Genes *CT-HULK1*, *CT-HULK2*, *CT-HULK3, PRP40,* and *AtCTD*

Table 2.2. List of genes/alleles cloned into yeast two-hybrid vector for yeast two hybrid assays

S.No	Gene/alleles	Cloning vector	Yeast-two hybrid vector	Construct name	Source
1	*HUA2*		pDBLeu	HUA2pDBLeu	Uday Sajja
2	*HUA2*		pEXP-AD	HUA2pEXP-AD	Uday Sajja
3	*HUA2*		pGBD	HUA2pGBD	Scott Deroo
4	*HUA2-ΔRPR*		pDBLeu	HUA2-ΔRPRpDBLeu	Uday Sajja
5	*HUA2-ΔRPR*		pEXP-AD	HUA2-ΔRPRpEXP-AD	Uday Sajja
6	*hua2-1*		pDBLeu	hua2-1pDBLeu	Uday Sajja
7	*CT-hua2*		pDBLeu	CT-HUA2pDBLeu	Uday Challa
8	*HULK1*		pEXP-AD	HULK1pEXP-AD	Sathya Challa
9	*HULK2*		pEXP-AD	HULK2pEXP-AD	Sathya Challa
10	*HULK3*		pEXP-AD	HULK3pEXP-AD	Sathya Challa
11	*FCA*		pEXP-AD	FCApEXP-AD	Uday sajja
12	*AtCTD*		pGAD	AtCTDpGAD	Dr. Culbertson
13	*AtCTD S2S5*		pGAD	AtCTD S2S5pGAD	Dr. Culbertson
14	*AtCTD A2S5*		pGAD	AtCTD A2S5pGAD	Dr. Culbertson
15	*AtCTD S2A5*		pGAD	AtCTD S2A5pGAd	Dr. Culbertson
16	*hua2-5*	pBSK	pDBLeu	hua2-5pDBLeu	
17	*CT-hulk1*		pDBLeu	CT-hulk1pDBLeu	
18	*CT-hulk2*		pDBLeu	CT-hulk2pDBLeu	
19	*CT-hulk3*		pDBLeu	CT-hulk3pDBLeu	
20	*PRP40*		pEXP-AD	PRP40pEXP-AD	
21	*PRP40(ww)*	pGEMT Easy	pEXP-AD	ww-PRP40pEXP-AD	
22	*RBP45*		pEXP-AD	RBP45pEXP-AD	
23	*UBP1*		pEXP-AD	UBP1pEXP-AD	
24	*AtCTD*		pDBLeu	AtCTDpEXP-AD	
25	*AtCTD*		pEXP-AD	AtCTDpEXP-AD	

Entries printed in red were cloned by previous lab members or were obtained from an outside research groups

were amplified from Col-0 cDNA using the indicated primers (Table 2.3) that incorporated suitable restriction sites. The resulting PCR products were digested with the appropriate restriction enzymes and were directly cloned into either pDBLeu or pEXP-AD. All vector inserts were sequenced to verify in frame fusion with the GAL4 DNA binding domain and/or GAL4 activation domain.

2.5.2 Cloning of genes and alleles for GST pull-down assay

Table 2.4 lists the genes cloned into bacterial expression vectors used for GST pull-down assay. For the expression of full-length HUA2, full length *HUA2* cDNA was cloned into two different vectors, pDEST™17 Vector (Invitrogen) and pET15b (Novagen). For cloning into pDEST17, full length *HUA2* cDNA was amplified using primers having *att*B sites. The PCR product flanked by *attB1* and *attB2* sequences was cloned into pDONOR^{TM}221 (Invitrogen) to get HUA2pENTRY using GATEWAY™ BP Clonase™ (Invitrogen). *HUA2* was then transferred into Gateway® pDEST™17 Vector (Invitrogen) to produce HUA2pDEST17 to produce a fusion protein of HIS tagged HUA2 protein.

For cloning into pET15b, full-length *HUA2* cDNA was amplified using primer pairs of HUA2pET15b Fwd/Rev that incorporated suitable restriction sites (Listed in Table 2.5). The PCR product was digested with the appropriate restriction enzymes and cloned directly into pET15b (Novagen). This produced constructs HUA2pET15b. The vector insert was verified by sequencing to ensure an in frame fusion to the HIS tag.

For producing constructs carrying the carboxyl-terminal (CT) ends of HUA2, HULK1, HULK2, and HULK3 the cDNAs of the amino-terminus end of *HUA2*, *HULK1*, *HULK2*, and *HULK3* were amplified and cloned directly into pET15b using the indicated primer

Table 2.3. List of primers used for cloning genes into yeast expression vector for yeast two-hybrid assay

Primer Name	Restriction enzyme	Sequence (FWD, REV)
hua2-5 pDB		GGGAATGGCTCCTGGGCGAAAAGAGG,
	NotI	GCGGCCGCATGGTATCTGATGATGGCTAG
PRP40(ww) pDB	*SalI*	CGGGTCGACTAGTCAGGCACAACAGAGAGC,
	NotI	TACGCGGCCGCGGAAGTTCTTCACCG
PRP40 pDB	*SmaI*	CGACCCGGGTATGGCTAATAACCATC,
	NotI	GGAGCGGCCGCTCAATACCGACCAC
PRP40 pEXP	*SmaI*	CGACCCGGGTATGGCTAATAACCATC,
	NotI	GGAGCGGCCGCTCAATACCGACCAC
AtCTDpDB	*SalI*	GGCTTGTCGACCCCAATGTCAGATGCACAG,
	NotI	GCGTGCGGCCGCTCAAGGGTTGCCTTTATC
AtCTDpEXP	*SalI*	GGCTTGTCGACCCCAATGTCAGATGCACAG,
	NotI	GCGTGCGGCCGCTCAAGGGTTGCCTTTATC
UBP1pEXP	*SalI*	GTCGACAATGCAGAGGTTGAAGCAGCAGCAG,
	NotI	**GCGGCCGC**TTACTGGTAGTACATGAGCTGCTG
CT-HULK1 pDB	*SalI*	GTCGACCATGGAAGAGCCAGTCCTTGC,
	NotI	GCGGCCGCTTTATAACTCTCTCTCAAGTTTCT
CT-HULK2 pDB	*SalI*	GTCGACTGTCTCG TCATCCACGGCTG,
	NotI	GCGGCCGCTTTATAACTCTCTCTCAAGTTTCT
CT-HULK3 pDB	*SalI*	GTCGACCGAAGGAAGTGACTCGGATGG,
	NotI	GCGGCCGCTCTTCATCTCTTGTCTCCCCAC

Nucleotide bases in blue represent extra bases added for restriction digestion. Nucleotide bases in red represent restriction sites

Table 2.4. List of genes cloned into bacterial expression vector for GST pull-down assay

S.No	Gene	Cloning vector	Expression vector	Construct name
1	*HUA2*	pENTRY	pDEST17	HUA2pENTRY
2	*HUA2*		pET15b	HUA2pET15b
3	*CT-HU2*		pET15b	CT-HU2pET15b
4	*CT-HULK1*		pET15b	CT-HULK1pET15b
5	*CT-HULK2*		pET15b	CT-HULK2pET15b
6	*CT-HULK3*		pET15b	CT-HULK3pEt15b
7	*PRP40(ww)*		pGEX-4T1	PRP40(ww)pGEX4T
8	*RBP45*		pGEX-4T1	RBP45pGEX4T
9	*UBP1*		pGEX-4T1	UBP1pGEX4T
10	*ww-FCA*		ww-FCA	ww-FCApGEX4T
11	*FCA*		pGEX-4T1	FCApGEX4T
12	*AtCTD*		pGEX-4T1	AtCTDpGEX4T

Table 2.5. List of primers used for cloning genes into bacterial expression vector for GST pull-down assays

Primer Name	Restriction enzyme	Sequence (Fwd, Rev)
HUA2pET15b	*NdeI* *BlpI*	GGTCCATATGATGGCTCCTGGGCGTAAAAG, TACCGCTCAGCTCAAGCTGGCCTCCAAC
CT-HUA2 pET15b	*NdeI* *BamHI*	CGCCATATGGATAAATGCCATCGAGTTTTGGAG, CGCGGATCCTCAAGCTGGCCTCCAACAATTCACG
CT-HULK1 pET15b	*XhoI* *BamHI*	GCGCTCGAGATGGAAGAGCCAGTCCTTG, CGCGGATCCTCAGGATGGCCTGTACTG
CT-HULK2 pET15b	*XhoI* *BamHI*	GCGCTCGAGGTCTCGTCATCCACGGCTG, CGCGGATCCCTAGTCACTTCTCTGATG
CT-HULK3 pET15b	*XhoI* *BamHI*	GCGCTCGAGGAAGGAAGTGACTCGGATG, CGCGGATCCTCACCTTTGATGCCACGTCCC
ww-FCA pGEX4t	*SalI* *NotI*	CCGGTCGACTAGGCACCTCCTGTTGGACTTGG, GCCGCGGCCGCTCAAGCTTTATTCTTCC
FCA pGEX4t	*SalI* *NotI*	CGAGGTCGACCAATGAATGGTCCCCCAG, TTAGCGGCCGCTCAAGCTTTATTCTTCC
PRP40(ww) pGEXT4t	*SalI* *NotI*	GCGGGTCGACGTCAGGCACAACAGAGAG, GTAGCGGCCGCCACTACCCTGAGGGACAG
RBP45 pGEX4t	*SalI* *NotI*	CCGGTCGACTCATGATGCAGCAGCCACCACC, GATGCGGCCGCTTAGTAGCTAAACCCGAC
UBP1 pGEXT4t	*SalI* *NotI*	CCGGGTCGACAAATGCAGAGGTTGAAGC , GTAGCGGCCGCTTACTGGTAGTACATGAGCT
AtCTD pGEX4t	*EcoRI* *EcoRI*	TGCAGAATTCCAATGTCAGAT, ACGTGAATTCAAGGGTTGCCT

Nucleotide bases in blue represent extra bases added for restriction digestion. Nucleotide bases in red represent restriction sites

pairs for each of the specific genes (Table 2.5). The same strategy used for cloning full-length *HUA2* cDNA into pET15 was used to produce constructs CT-HUA2pET15b, CT-HULK1pET15b, CT-HULK2pET15b and CT-HULK3pET15b. Vector inserts were sequenced to verify in frame fusion to the HIS tag.

The cDNA of *AtCTD, FCA, FCA*(*ww*), *PRP40(ww), RBP45*, and *UBP1* were directly cloned individually into pGEX4T1 (Amersham Biosciences) using the appropriate primer pairs (Table 2.5) for each specific gene. This produced constructs AtCTDpGEX4T1, FCApGEX4T1, WW-FCApGEX4T, PRP40(ww)pGEX4T1, RBP45 pGEX4T1 and UBP1pGEX4T1. Vector inserts were sequenced to verify in frame fusion to the GST tag.

2.5.3 Cloning of genes for *In planta* assays

Constructs for the *in planta* pull-down assay and BiFC assay using plant protoplast and tobacco infiltration are listed in Table 2.6. All of the constructs for these assays were generated by Dr. Sascha Laubinger, The Center for Plant Molecular Biology (ZMBP) University of Tübingen, Germany.

2.5.4 Cloning of genes for intracellular localization assay

Constructs for the intracellular localization assay using particle bombardment are listed in Table 2.7. The cDNA of *hua2-1* allele and the amino terminus region of *HUA2 (NT-HUA2)* carrying the nuclear localization signals was amplified from HUA2pDBLeu (Sajja Ph.D. thesis, 2009) using the indicated primer pairs (Table 2.8). The cDNAs of *hua2-1* and *NT-HUA2* were first cloned into pGEM®-T Easy (Promega) Vector according to the manufacturer's instructions.

Table 2.6 List of genes cloned into *in planta* BiFC vector for BiFC assay in plant protoplast.

S.No	**Gene/alleles**	**Cloning vector**	**BiFC vector**	**Construct name**
1	*HUA2*	pENTRY	pARC235	HUA2pARC235
2	*CT-HUA2*	pENTRY	pARC233	CTHUA2pARC233
3	*ww-FCA*	pENTRY	pARC234	wwFCApARC234
4	*WF-FCA*	pENTRY	pARC234	WF-FCApARC234
5	*RBP45*	pENTRY	pARC234	RBP45pARC234
6	*UBP1*	pENTRY	pARC234	UBP1pARC234

Table 2.7. List of genes cloned for intracellular localization assays

S.No	Gene/alleles	Cloning vector	Binary vector	Construct name
1	*HUA2*	pDONOR221	pEarleyGate 103	HUA2pEARLY103
2	*HULK1*	pCR8-TopoGW	pBIN19	HULK1pBIN19
3	*HULK2*	pCR8-TopoGW	pBIN19	HULK2pBIN19
4	*HULK3*	pCR8-TopoGW	pBIN19	HULK3pBIN19
5	*CT-HUA2*	pCR8-TopoGW	pEarleyGate 104	CTHUA2pEARLY104
6	*NT-HUA2*	pGEMT-Easy	pEZS-NL	NT-HUA2pEZ-NL
7	*NT-HUA2*	pGEMT-Easy	pEZS-CL	NT-HUA2pEZ-CL
8	*hua2-1*	pGEMT-Easy	pEZ-NL	hua2-1pEZ-NL
9	*hua2-1*	pGEMT-Easy	pEZ-CL	hua2-1pEZ-CL

Table 2.8. List of primers used for cloning genes into vectors for intracellular localization assays

Primer Name	Restriction enzyme	Sequence(FWD, REV)
NT-HUA2 pEZ-NL/PJGFP1	*XhoI*	CTCGAGATGGCTCCTGGGCGTAAAAG,,
PJGFP6	*SmaI*	CCCGGGATGCATCCTCACAATGTTTACG
NT-HUA2 pEZ-CL/ PJGFP2	*KpnI*	GGTACCCTATGGCTCCTGGGCGTAAAAGAGG,
PJGFP5	*SmaI*	CCCGGGTGCATCCTCACAATGTTTACG
hua2-1 pEZ-NL/PJGFP1	*XhoI*	CTCGAGATGGCTCCTGGGCGTAAAAG,
PJGFP4	*SmaI*	CCCGGGTGCCATACTCATCAACAAGC
hua2-1 pEZ-CL/PJGFP2	*KpnI*	GGTACCCTATGGCTCCTGGGCGTAAAAGAGG,
PJGFP3	*SmaI*	CCCGGGGCCATACTCATCAACAAGC
HUA2		GGGGACAAGTTTGTACAAAAAAGCAGGCTT CATGGCTCCTGGGCGTAAAAGAG, GGGGACCACTTTGTACAAGAAAGCTGGGTCA GCTGGCCTCCAACAATTCAC
CT-HUA2		ATGGCCTCGAGTGATAAATG, TTCCTCATTTCTGTATCGCC

Nucleotide bases in red represent restriction sites

The cDNA was then sub-cloned into two different vectors, pEZ-NL and pEZ-CL. For sub-cloning into pEZ-NL, *hua2-1* and *NT- HUA2* were first digested from pGEMT easy using *XhoI* and *SmaI* restriction enzyme and later ligated into the *XhoI* and *SmaI* sites of pEZ-NL to create hua2-1pEZ-NL and NT-HUA2pEZ-NL. For sub-cloning into pEZ-CL, *hua2-1* and *NT-HUA2* were first digested from pGEMT easy using *KpnI* and *SmaI* restriction enzyme and later ligated into the *KpnI* and *SmaI* sites of pEZ-NL to create hua2-1pEZ-CL and NT-HUA2pEZ-CL.

Full length *HUA2* and the carboxyl-terminal region of *HUA2 (CT-HUA2)* were cloned into pEarleyGate103 and pEarleyGate104respectively, to create HUA2pEarley103 and CT-HUA2pEarley104. For producing HUA2pEarley103, full length *HUA2* cDNA was amplified using primers incorporating *att*B sites (Listed in Table 2.8). The PCR product flanked by*attB1* and *attB2* sequences, was cloned into pDONOR^{TM}221 (Invitrogen) to get HUA2pENTRY using GATEWAY™ BP Clonase™ (Invitrogen). *HUA2* was then transferred into pEarleyGate 103 (35S-Gateway-GFP-His tag-OCS 3') (www.arabidopsis.org, ABRC stock number CD3 685) using GATEWAY™ LR Clonase™ (Invitrogen) to create binary HUA2pEarley103 expression vector. For producing CT-HUA2pEarley104, the cDNA region of carboxyl-terminal end of *HUA2* was amplified using primers incorporating *attB* sites (Table 2.8). The PCR product flanked by *attB1* and *attB2* sequences was cloned into pCR8-Topo-GW (Invitrogen) to get CT-HUA2pENTRY using GATEWAY™ BP Clonase™ (Invitrogen). *CT-HUA2* was then transferred into the pEarleyGate 104 (35S-YFP-Gateway-OCS 3') (www.arabidopsis.org, ABRC stock number CD3 686) using GATEWAY™ LR Clonase™ (Invitrogen) to create binary CT-HUA2pEarley104 expression vector.

2.5.5 Cloning of genes for domain characterization assay

The constructs created and used for the domain characterization assay are listed in Table 2.9. HUA2pEarley100 was produced by performing an LR reaction with HUA2pENTRY and pEarleyGate 100 using GATEWAY™ LR Clonase™ (Invitrogen). ΔPWWP-HUA2 pEarley100 construct carries *HUA2* cDNA which express HUA2 with mutated PWWP domain. The conserved Pro-Trp-Trp-Pro motif within the PWWP domain was replaced with Ala-Ala-Ala-Ala. *ΔPWWP-HUA2* cDNA, was amplified from HUA2pENTRY using primers HUA2-PWWP-FWD with *ApaI* restriction site and SDM-PWWP-REV with *NgoMIV* restriction site (Listed in Table 2.10). SDM-PWWP-REV primer was designed such that it introduces 4 alanines (AAAA) in place of PWWP amino acids. The amplified product was later digested with *ApaI* and *NgoMIV* and sub-cloned into HUA2pENTRY between *ApaI* and *NgoMIV* to produce ΔPWWP-HUA2pENTRY. The sequence of *ΔPWWP-HUA2* cDNA was sequenced to confirm the desired change within PWWP motif was made. Later a LR reaction was performed between ΔPWWP-HUA2pENTRY and pEarleyGate 100 using GATEWAY™ LR Clonase™ (Invitrogen) to produce ΔPWWP-HUA2 pEarley100.

Table 2.9. List of genes cloned for domain characterization assays

S.No	**Gene/alleles**	**Cloning vector**	**Binary vector**	**Construct name**	**Source**
1	*HUA2*	pGEMT-Easy	pCAMBIA	HUA2pCAMBIA	Uday sajja
2	*HUA2*	pGEMT-Easy	pCAMBIA	ΔRPR-HUA2pCAMBIA	Uday sajja
3	*HUA2*	pDONOR221	pEARLY100	HUA2pEARLY100	
4	*HUA2*	pDONOR221	pEARLY100	ΔPWWP-HUA2pCAMBIA	

Table 2.10. List of primers used for cloning genes into binary vectors for domain characterization assays

Primer Name	Restriction enzyme	Sequence (FWD,REV)
HUA2-PWWP	ApaI	GCTCGGGCCCCAAATAATG
	NgoMIV	CTTCAGGCCGGCTGATCTTGGCAGCCGCAGCAGCGAAGCCTTTGAC

Nucleotide bases in red represent restriction sites

2.6 Yeast two-hybrid assay

Yeast two-hybrid assays were performed according to the protocol described in the ProQuest™Two-Hybrid System instruction manual (Invitrogen). Yeast transformation was carried out as described earlier (Section 2.4.6). Before performing the assay, the levels of auto-activation of the bait (GAL4 DB fusion) proteins were analyzed for nonspecific according to the manufacture's protocol. Two different yeast strains were used for the yeast two-hybrid assay, MaV203 and PJ69-4A. For MaV203, vectors pDBLeu and pEXP-AD502 were used. Gene/alleles cloned into pDBleu were selected on YNB-Leucin amino acid dropout media. Genes/alleles cloned into pEXP-AD502 were selected on YNB-Tryptophan amino acid dropout media. For strain PJ69-4A, vectors GBD-C1 and GAD-C1 (kindly provided by Dr. Culbertson, University of Wisconsin, Madison, Wisconsin) were used. For both strains, *HIS*3 acted as the reporter gene for a positive interaction.

Refer Appendix I Tables A1 to A8 for the combination of fusion proteins with GAL4 DB and GAL4 AD for which interaction was checked in yeast strain MaV203. Refer Appendix I Table B1 for the combination of fusion proteins with GAL4 DB and GAL4 AD for which interaction was checked in yeast strain PJ69-4A.

2.7 GST pull-down assay

2.7.1 Expression of GST tagged bait proteins and HIS tagged prey proteins

To express GST alone (control) and GST tagged WW-RPP40, WW-FCA, FCA, RBP45, UBP1 and AtCTD and HIS tagged HUA2, CT-HUA2, CT-HULK1, CT-HULK2 and CT-HULK2, the respective vectors were transformed into *E. coli* strain BL21 (DE3) pLysS

(Novagen). A single colony was picked and inoculated in 5 ml of LB medium containing 50 μg/ml ampicillin (final concentration) and incubated at 37 ^{o}C in a shaking incubator for starter culture. The starter culture was added to a flask containing 500 ml of fresh LB plus ampicillin and incubated in a 37 ^{o}C shaker with vigorous shaking for over two hours until the cells reached mid-log growth (A_{550}=0.5–1.0). At this time, 1 mM isopropylthio-β-galactoside (IPTG) was added to the culture to induce expression of the fusion bait/prey protein. The flask was transferred to a 30 ^{o}C shaking incubator and incubated for an additional 6 hours. Next the liquid culture was centrifuged at 7,000x g for 10 minutes to pellet cells. Later the bacterial cell pellet was resuspended in 10 ml lysis buffer (1x PBS pH 7.4 with 1% TritonX-100, 1 Mm PMSF and Complete Protease Inhibitor Cocktail (Roche). Resuspended cells were lysed by sonication on ice for 180 sec (6 times, 30 sec). The supernatant was cleared of cell debris by centrifugation at 15,000x g for 20 min at 4^{o}C. Protein concentration was determined by the Bradford method (Bio-Rad Ltd., Canada). Approximately 3 mg of total cellular lysate containing GST-fusion bait protein and GST alone (control) were incubated in two different microfuge tubes having 125 μl (bed volume) of pre-washed Glutathione-Sepharose 4B beads (GE healthcare) in a cold room (4^{o}C) on a rotary shaker for 2 hours. The microfuge tube having Glutathione-Sepharose 4B beads immobilized with GST-bait fusion and GST protein was washed once with PBS buffer containing 500 mM NaCl, followed by 5 times washing with PBS buffer containing 135 mM NaCl to remove all non-specific and unbound protein from the column. Next 5 mg of total cellular lysate containing HIS-fusion prey protein was loaded into tubes containing immobilized GST-bait protein (test) and GST alone (control). Simultaneously, to check the binding of HIS-prey protein to

Glutathione-Sepharose 4B beads, an equal amount (5 mg) of total cellular lysate containing HIS-fusion prey protein was loaded onto a microfuge tube containing 125 µl (bed volume) of pre-washed Glutathione-Sepharose 4B beads. The microfuge tubes were left on a rotary shaker overnight in a cold room (4^{o}C). The following day all the three tubes (one test and two controls) were washed once with 500 µl PBS buffer containing 135 mM NaCl and three times with 500 µl PBS buffer containing 50 mM NaCl. After the washes, the protein complex retained on the beads was extracted into 50 µl of 2X SDS-sample buffer by boiling for 5 min, separated by SDS-PAGE and immunoblotted with HIS and GST antibody.

2.9 Detection of proteins by western blotting

Proteins were separated on 7.5% (w/v) SDS-polyacrylamide gels and transferred onto nitrocellulose membranes by electro-blotting using a Trans-Blot Semi-Dry Electrophoretic Transfer Cell (Bio-Rad). Membranes were blocked overnight at 4 ^{o}C with 1X PBST buffer containing 5% fat-free milk powder. To check for presence of HIS-prey protein, the blot was probed with HIS antibody (GenScript) (5 mg/ml) at a dilution of 1:10,000 (v/v) for 2 hours. To check for presence of GST-bait proteins, the blot was probed with GST antibody (GenScript) (5 mg/ml) at a dilution of 0.5:10,000 (v/v) for 1 hour. Membranes were then incubated with peroxidase-labeled goat anti-mouse IgG (0.5:10,000) for 90 min. The antigen-antibody complex was detected by chemiluminescence using SuperSignal West Pico Chemiluminescent Substrate kit (Pierce). For the GST pull-down assay, the membrane was first probed with HIS antibody, stripped with Restore Western Blot stripping buffer (Pierce) and probed with GST antibody.

2.10 Intracellular localization assay

2.10.1 Particle bombardment

Allium cepa cv. Tango bulbs were purchased from local grocery stores. Under sterile conditions the inner epidermal peels were removed from the bulb and placed on petri plates with 1x Murashige and Skoog (MS) salts having 30 g/L sucrose and 2% agar. 5 µl (1 µg/µl) of constructs encoding GFP-NT-HUA2, NT-HUA2-GFP, GFP-hua2-1 and hua2-1GFP were coated onto tungsten microcarriers according to the manufacturer's instructions (Biolistic PDS-1000/He system, Bio-Rad). Particles were bombarded onto the onion epidermal cells using the Biolistic PDS-1000/He system with 1100 psi rupture discs under vacuum of 28 in Hg. After bombardment, the cells were allowed to recover in darkness for 24 hours on agar plates at 22 °C. To visualize GFP signals, those cells bombarded with the constructs were observed under a fluorescence microscope.

2.10.2 Intracellular localization assay using agro-infiltration into *Nicotiana benthamiana* leaves

Appropriate constructs for the assay was transformed into agrobacterium cells strain GV3101 using electroporation (refer to Section 2.4.4). The colonies used for tobacco infiltration were checked by colony PCR to confirm the presence of plasmid/gene. A single colony of this was inoculated in 30 ml YEB medium and incubated overnight under shaking at 28 °C. These cultures were centrifuged at 4,000x g for 15 minutes and the resulting pellet was resuspended in 1 ml of AS medium (AS-medium, 100 ml: 1 ml 1 M MES-KOH, pH 5.6, 333 µl of 3 M $MgCl_2$, 100 µl of 150 mM acetosyringone in DMSO). The samples were later diluted with AS medium to achieve an OD_{600}=0.7–0.8

(approximately 1 ml per leaf). For a positive control a plasmid encoding mCherry-NLS (nuclear marker) under the control of a constitutive promoter was used. The positive nuclear marker was derived from the vector described in Karimi et al. (2005) and Shaner et al. (2004), and was obtained from Prof. Dr. Klaus Harter, ZMBP, The University of Tuebingen. The p19 protein from tomato bushy stunt virus cloned in pBIN61 (Voinnet et al., 2000) was used to suppress gene silencing. A working suspension was prepared by mixing *Agrobacterium* strains carrying each of the following individual constructs: HUA2pEARLEY103, HULK1pBIN19, HULK2pBIN19, HULK3pBIN19 and CT-HUA2pEARLEY104 with the p19 plasmid and a plasmid encoding mCherry-NLS at a ratio of 1:1:1. The mixed *Agrobacterium* strains were co-infiltrated into the total abaxial air space of tobacco leaves using a 1 ml syringe. The infiltrated plants were transferred to a greenhouse and after 2–3 days the abaxial epidermis of infiltrated tobacco leaves were checked for fluorescence by Confocal laser-scanning microscopy (CLSM) using a Leica TCS SP2 confocal microscope (Leica Microsystems GmbH).

2.11 *Agrobacterium*-mediated transformation of *Arabidopsis thaliana FRI hua2-3* and *hua2-7/hulk1 plants* using the floral dip method

For RPR and PWWP domain characterization experiments, agrobacterium strains carrying individual constructs of pEarleyGate 100, HUA2pEarley100, ΔPWWP-HUA2pEarley100, pCAMBIA, HUA2pCAMBIA and ΔRPR-HUA2pCAMBIA were individually transformed into *FRI hua2-3* and *hua2-7/hulk1 plants* using the floral dip method (Zhang et al., 2006). After transformation, dried T_1 seeds were collected from primary (T_0) plants and screened for putative transformants. For T_1 seeds carrying pEarleyGate 100, HUA2pEarley100, ΔPWWP-HUA2pEarley100, BASTA (250 µg/ml)

selection was carried out on soil for putative transformation. For T_1 seeds carrying pCAMBIA, HUA2pCAMBIA and ΔRPR-HUA2pCAMBIA putative transformants were selected on MS media (Sigma) having hygromycin at a concentration of 30 µg/ml and 25 µg/ml of carbenicillin. Putative transformants from both soil and agar plates were transplanted into fresh pots with soil (PRO MIX, plant products) under 100-150 µmol photons/m^2 /s cool-white light at 23 °C. Since the number of basal rosette leaves formed is directly correlated with flowering time, the flowering time was scored as the total number of basal rosettes when the inflorescence stem had reached 1-cm height for both *FRI hua2-2* and *hua2-7/hulk1* T_1 plants. For the T_1 hua2-7/hulk1 plants, the embryo phenotype was examined by split opening the mature green siliques. We observed 5 siliques from each of the 10 plants for each construct under the dissecting microscope. The number of defective and wild-type embryos was recorded for each silique. The number of wild-type and defective embryos was averaged for each plant.

2.12 Statistical Analysis

Data were statistically analysed to determine if the flowering time and embryo defect were rescued using SAS v. 9.2 (SAS Institute, Inc. 2003). One-way ANOVA was used to test for significant differences in basal leaf number (flowering time) and mean number of wildtype and defective embryos (embryo phenotype) in the three constructs for each experiment. A significance value of 0.05 was used for alpha and all data were tested for normality and homogeneity of variance prior to the ANOVA. For the multiple comparison procedure after a significant ANOVA, the least squares means were calculated using the LSMEANS statement in SAS and means for each construct were compared with Student's t-test using the PDIFF statement. Alpha was manually adjusted

for each comparison using the Bonferroni correction, where alpha is divided by the total number of comparisons made.

Chapter 3: Results

3.1 HUA2 physically interacts with RBP45 and UBP1 in yeast cells

HUA2 was first identified in a genetic screen aimed at isolating mutants that enhance the defective flower development phenotype displayed by the weak *AG* allele, *ag-4*(Chen and Meyerowitz, 1999). *HUA2* affects splicing and 3` end processing of *AG* and *FCA*(Cheng et al., 2003; Wang, Sajja and Grbic unpublished) and is also required for the proper transcriptional regulation of the *FLC* gene (Doyle et al., 2005; Wang et al., 2007). Interestingly, genetic analysis performed on *Arabidopsis* accession *Sy-0* indicates that *HUA2*'s function is also important for the morphological properties characteristic of the *Sy-0* accession (Poduska et al, 2003).

Though *HUA2* is known to modulate plant developmental processes, it is unclear how it functions at the molecular level. The domain architecture of HUA2 suggests that it may have roles in chromatin targeting and/or RNA processing. In an attempt to understand the molecular mechanism by which *HUA2* acts on its targets, our laboratory utilized the yeast two-hybrid system to identify interacting partners of HUA2.

The two-hybrid assay identifies interactions between two proteins (X and Y) by reconstituting an active transcription factor (GAL4) within a yeast cell. The active transcription factor is formed by the reconstitution of two fusion proteins: one that contains the GAL4 **D**NA-**B**inding domain (DB) fused to the first protein of interest (DB-X; also known as the "bait"), and another that contains the GAL4 **A**ctivation **D**omain (AD) fused to the second protein of interest (AD-Y; also known as the "prey" or "target protein"). Interaction between DB-X:AD-Y reconstitutes the functional GAL4

transcription factor. The reconstituted GAL4 is then able to activate a reporter gene integrated into the chromosome of the host yeast strain. The reporter genes used in my analyses are *HIS3*, *URA3* and *lacZ*. In cases where *HIS3* or *URA3* were used as the reporter gene, two-hybrid-dependent transcription activation was monitored by growing cells on plates lacking histidine or uracil, respectively. In cases where the *lacZ* reporter was used, transcriptional activation was determined by a colorimetric assay; in the presence of X-gal yeast colonies expressing *lacZ* turn blue in color.

Using HUA2 as bait, a yeast two-hybrid screen was performed using a library of AD fusions derived from a vegetative meristem cDNA library of *Arabidopsis*. After three rounds of sub-screening, a total of seven putative positive clones were identified (listed in Table 3.1; Sajja, 2009). Among the putative interactors identified in the screen, two were methyl-CPG-binding domain proteins (1 and 4), one transcription factor (SPL9), two splicing factors (UBP1 and RBP45) one cyclin dependent protein kinase and one clathrin binding protein. UBP1 and RBP45 were of particular interest since our previous data indicated that HUA2 may be involved in the processing of pre mRNAs (Cheng et al., 2003, Sajja, 2009). Interestingly HUA2 has also been shown to participate in splicing and has also been shown to interact with PRP40 and FCA; two other splicing factors (Sajja 2009 and Cheng 2003). This suggests that HUA2 might act as part of a complex with these proteins to regulate pre-mRNA processing. To better understand the functional significance of HUA2 domain structure with respect to its interaction with FCA, PRP40, RBP45 and UBP1, I undertook a series of experiments aimed at better defining the network environment of HUA2.

Table 3.1. Putative HUA2 interactors identified through yeast two-hybrid screening.

Protein name	Function	AGI Codes
AtMBD1	methyl-CPG-binding protein	At4g22745
AtMBD4	methyl-CPG-binding protein	At4g63030
SPL9	Transcription factor	At2g42200
RBP45	Splicing factor	At1g11650
UBP1	Splicing factor	At1g17370
PRP40[1]	Splicing factor	At3g19670
FCA[2]	Splicing factor	At4g16280
CYCP4	Cyclin dependent protein kinase	At2g44740
SH3P3	Clathrin binding	At4g18060

[1] identified in different screen (Cheng, 2003), [2] Identified in an independent assay (Sajja 2009)

To begin, the interaction between HUA2 and RBP45 and between HUA2 and UBP1 (identified in the genome-wide screen described above) were confirmed by performing individual yeast two-hybrid assays. To this end, MVa203 yeast cells were co-transformed with a plasmid carrying the *GAL4* DNA-binding domain fused to *HUA2* (GAL4 DB-HUA2) and with a plasmid containing the *GAL4*- activation domain fused to either, *RBP45* or *UBP1* (GAL4 AD-RBP45, GAL4 AD-UBP1*).* Activation of *HIS3* transcription was assayed by monitoring growth on medium lacking histidine. FCA and PRP40, which have been shown to interact with HUA2 in independent yeast two-hybrid assays, were used as positive controls (Sajja, 2009 and Cheng, 2003). Colony formation under stringent conditions (50mM 3-Amino-1, 2, 4-Triazole; 3AT) seen in Figure 3.1A, panels 1-4 indicates that both RBP45 and UBP1 interact with HUA2. No growth was seen in the negative control (Figure 3.1, lower panel). An X-gal filter assay, which measures induction of the *lacZ* reporter gene, resulted in blue colonies only for positive controls PRP40(WW) and FCA, but not for RBP45 and UBP1 (Figure 3.1B).As indicated by The ProQuest™ Two-Hybrid System manual, the X-gal filter assay is less sensitive than the growth assay and provides only a qualitative assessment of enzyme activity. The manual thus states to interpret interactions positive via growth assay and negative via X-gal assay to be considered weak interactors (Table 3.2). These data suggests that HUA2 shows a stronger preference for binding with the WW containing proteins.

Figure 3.1.Yeast two-hybrid analysis of the interaction between HUA2 and either PRP40(WW), FCA, RBP45 or UBP1. MVa203 cells were individually co-transformed with constructs encoding GAL4 DB-HUA2 and either A: (1) GAL4 AD-PRP40(WW), (2) GAL4 AD-FCA, (3) GAL4 AD-RBP45, (4) GAL4 AD-UBP1 or (5) GAL4 AD. Transformants were grown on selection plates with 50 mM 3-AT lacking leucine, tryptophan and histidine (-LTH). **B**: X-gal filter assays for the colonies shown in (**A**).

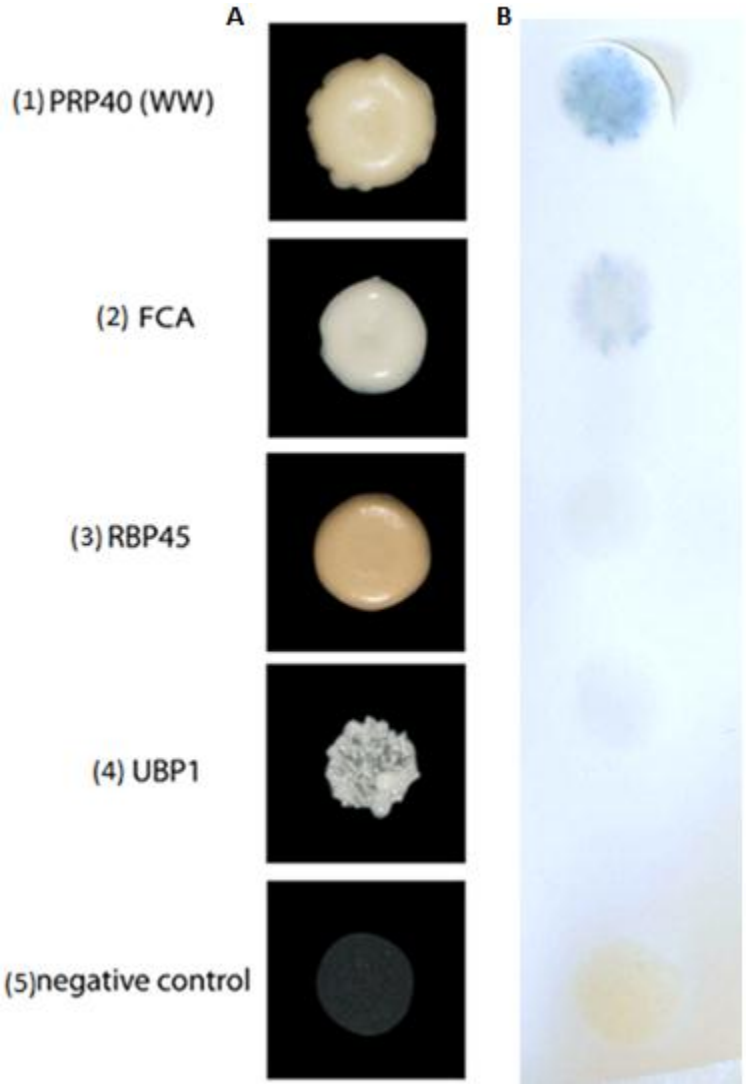

A
B
(1) PRP40 (WW)
(2) FCA
(3) RBP45
(4) UBP1
(5)negative control

Table 3.2. Summary and interpretation of yeast two-hybrid assays conducted with the indicated bait and prey proteins.

Fusion gene with				
*GAL*DB	*GAL*AD	*HIS*3 Expression	β-galactosidase Expression	Interpretation
	PRP40(WW)(P[1])	+	Dark Blue	Strong interaction
	FCA (P)	+	Blue	Strong interaction
HUA2	*RBP45*	+	White	Weak interaction [2]
	UBP1	+	White	Weak interaction
	Empty(N)	-	White	No interaction

[1]P indicates positive control and N indicates negative control.

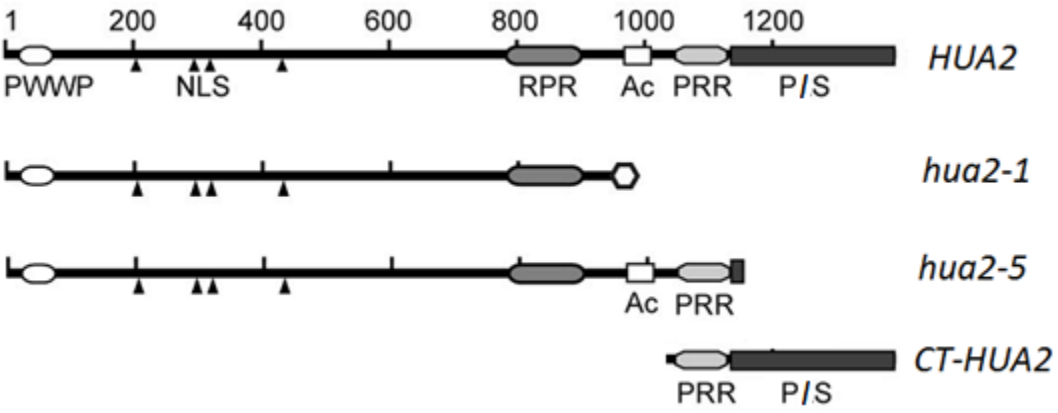
[2]Interactions positive via growth assay and negative via X-gal assay to be considered weak interactors.

3.2 The proline-rich sequence in the carboxyl-terminal region of HUA2 is required and sufficient for protein-protein interaction

The proline-rich sequence in the C-terminal region of HUA2 that contains the PPLP repeats has been suggested to be involved in mediating protein-protein interaction (Chen and Meyerowitz, 1999). It has also been shown that this proline-rich region is responsible for interaction with proteins containing the WW domain (Bedford et al., 1997). I thus decided to investigate the significance of these proline-rich sequences in the C-terminal region of HUA2 and their role in protein-protein interaction. To map the interacting domain within HUA2, I made a set of deletion constructs using the HUA2 alleles, *hua2-1* and *hua2-5*, both of which lack part of the C-terminal end of HUA2. In addition, a construct encoding only the C-terminal end of HUA2 was prepared (Figure 3.2). The *hua2-1* allele disrupts a splice junction at the sixth intron and produces a truncated protein lacks both the proline-rich region with PPLP repeats and the remainder of the C-terminal region rich in prolines/serines residues, (Chen and Meyerowitz, 1999). *hua2-5* contains a premature stop codon and produces a truncated protein which includes a PWWP domain, the four putative nuclear localization signals, the RPR domain and the proline rich region with the PPLP repeats. It lacks 247 amino acids at the entire C-terminus that includes a region rich in proline/serine residues (Doyle et al., 2005).

Constructs encoding GAL4 DB-HUA2, GAL4 DB-hua2-1 and GAL4 DB-CT-HUA2 were already available in the laboratory and their level of self-activation determined (Table 3.3). Thus, I created a construct encoding the GAL4 DB-hua2-5 fusion in the yeast two-hybrid vector pDBLeu. The self-activation of GAL4 DB-hua2-5 fusion was then checked to determine the basal levels of *HIS*3 reporter gene expression with bait

Figure 3.2. Domain structures of *HUA2* alleles used in yeast two-hybrid mapping assays. The *hua2-1* allele disrupts a splice junction at the sixth intron and produces a truncated protein retaining the proline-rich region with PPLP repeats but lacks the remainder of the C-terminal region rich in proline/serine residues. *hua2-5* contains a premature stop codon and produces a truncated protein containing the PWWP domain, the four nuclear localization ssignals, the RPR domain and the proline rich region with PPLP repeats, but lacks the 247 amino acid C-terminal that includes the region rich in proline/serine residues. *CT-HUA2,* C-terminal region of HUA2 (last 377 residues at C-terminal end) containing the proline rich region with PPLP repeats and the proline/serinerich residues.


1
200
400
600
800
1000
1200
PWWP
NLS
RPR
Ac
PRR
P/S
HUA2
hua2-1
hua2-5
Ac
PRR
CT-HUA2
PRR
P/S

alone. To check for self-activation, MVa203 cells were transformed with a plasmid encoding GAL4 DB-hua2-5. The transformed yeast cells were grown on plates lacking histidine and containing different concentrations of 3-Amino-1,2,4-triazole (3-AT). 3AT specifically inhibits the enzyme imidazoleglycerol phosphate dehydratase encoded by the reporter gene *HIS*3; thus any basal expression of the reporter gene can be "quenched" by the addition of 3AT and aid in preventing false positives. Plates lacking histidine and containing 40 mM of 3AT showed no growth of yeast cells, indicating that 40 mM 3AT is sufficient for complete inhibition of basal levels of the enzyme imidazole glycerol phosphate dehydratase (Figure 3.3). For all the yeast-two hybrid assays performed to check protein: protein interaction using *GAL*4 DB-*hua2-5*, a 3AT concentration of 50 mM was used. The 3AT concentrations required for inhibiting the self-activation of GAL4 DB-HUA2, GAL4 DB-hua2-1, GAL4 DB-hua2-5 and GAL4 DB-CT-HUA2 fusion proteins are listed in Table 3.3.

Next, I analysed whether the mutations in *hua2-1* and *hua2-5* affect the ability of HUA2 to bind with PRP40(WW), UBP1 or RBP45. In addition, I tested whether a C-terminal region of HUA2 (the C-terminal 377 residues) containing the proline rich region with the PPLP repeats along with the region rich in proline/serine is sufficient for interaction with PRP40(WW), UBP1 or RBP45. This was performed by individually co-transforming MVa203 cells with plasmids expressing the respective HUA2 fusion proteins together with plasmids expressing the GAL4 AD-PRP40(WW), GAL4 AD-UBP, or GAL4 AD-RBP45 fusions. The transformed cells were assayed for activation of *HIS3* and *Lac*Z transcription. Figure 3.4A shows that HUA2and CT-HUA2 but not hua2-1 or hua2-5 activated the *HIS*3 reporter gene under stringent conditions

Figure 3.3. Determination of the optimal concentration of 3AT for assaying GAL4 DB-hua2-5 fusion. The concentration of 3AT at which the expression of the *HIS*3 reporter gene was fully suppressed was determined by transforming MVa203 cells expressing *GAL4 DB-hua2-5* and plating them in quadruplicate on plates lacking leucine and histidine and containing the indicated concentration of 3AT.

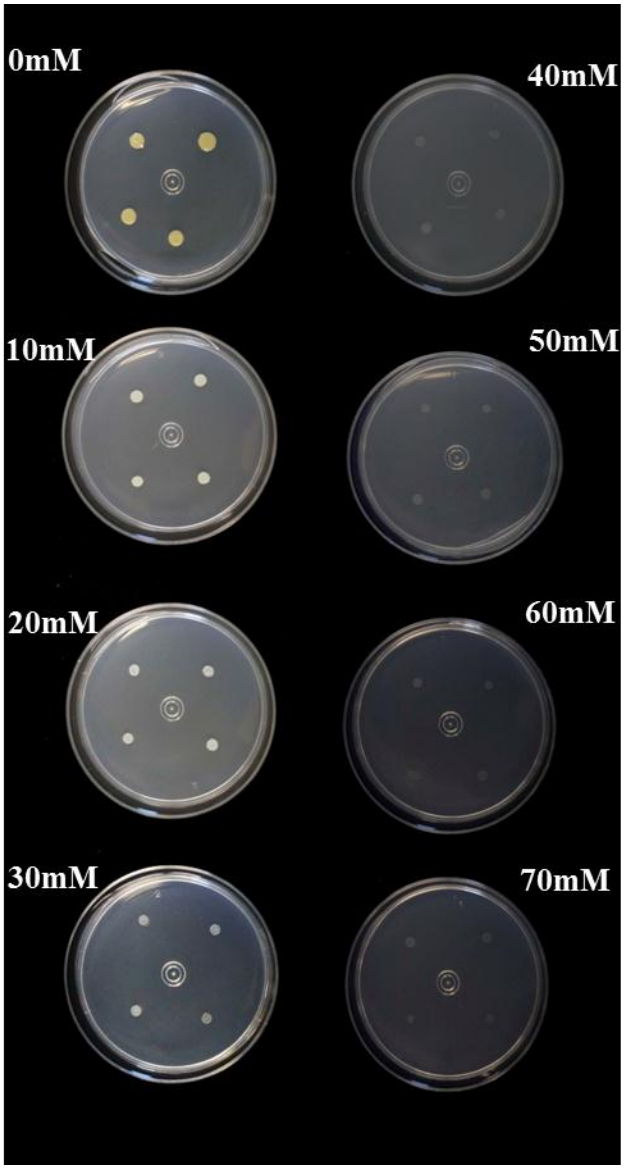
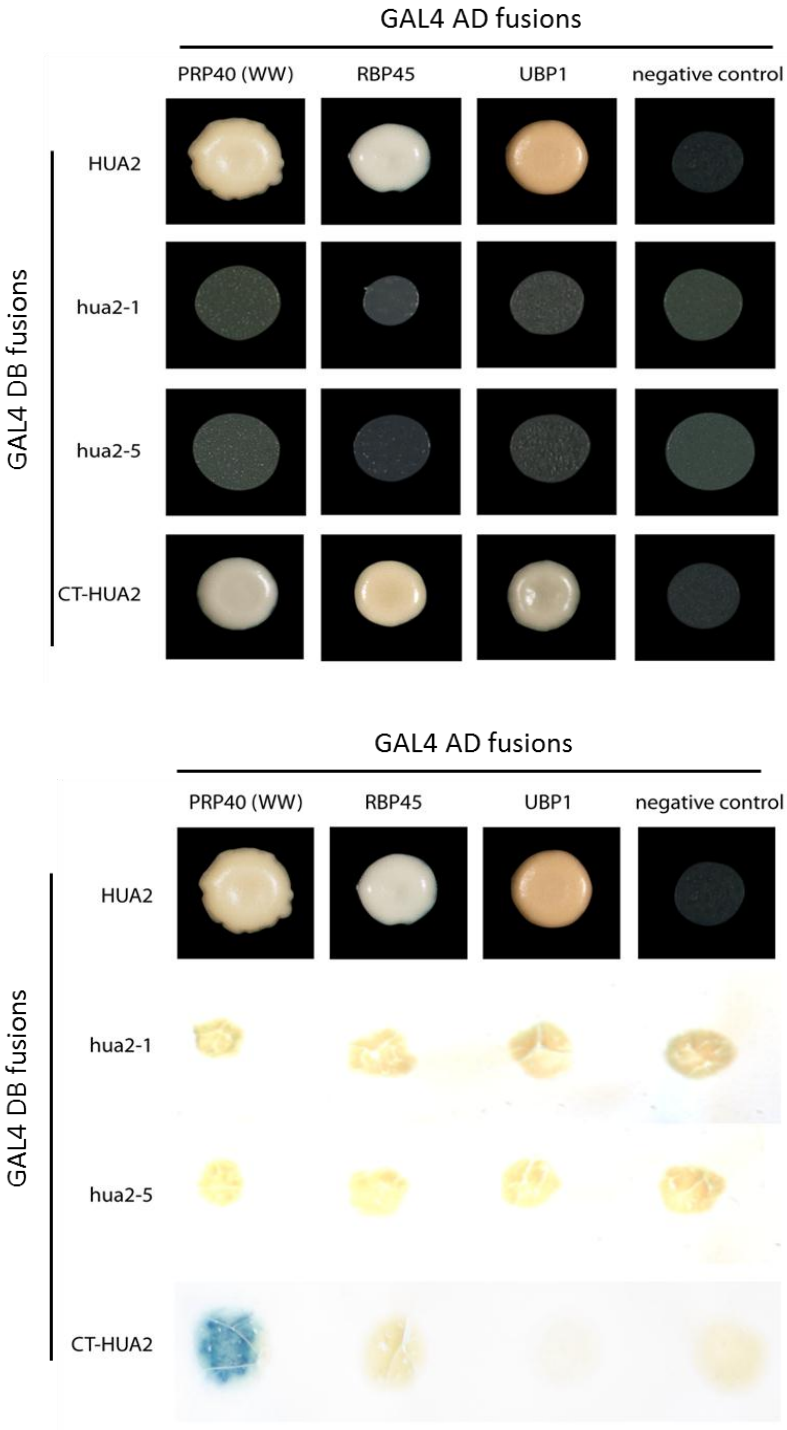

0mM
40mM
10mM
50mM
20mM
60mM
30mM
70mM

Table 3.3. 3AT concentrations required to inhbit self-activation of genes fused with *GAL4 DB*

Fusion gene with *GAL4 DB*	3 AT concentration required to inhibit self-activation
HUA2	50
hua2-1	50
hua2-5	50
CT-HUA2	80

Figure 3.4A.Yeast two-hybrid assay to map the interacting domain within HUA2. MVa203 cells were co-transformed with plasmids expressing the indicated *GAL4* DNA-binding domain and *GAL4*- activation domain fusions. *HIS3* reporter gene expression was assayed by growth on mediums lacking histidine.

Figure 3.4B.X-gal filter assay to map the interacting domain within HUA2. MVa203 cells were co-transformed with plasmids expressing the indicated *GAL4* DNA-binding domain and *GAL4*- activation domain fusions. *β-galactosidase* reporter gene expression was assayed by X-gal filter assay.


GAL4 AD fusions
PRP40 (WW)
RBP45
UBP1
negative control
GAL4 DB fusions
HUA2
hua2-1
hua2-5
CT-HUA2
GAL4 AD fusions
PRP40 (WW)
RBP45
UBP1
negative control
GAL4 DB fusions
HUA2
hua2-1
hua2-5
CT-HUA2

(50 mM 3AT for *HUA2, hua2-1, hua2-5* and 80 mM 3AT for *CT-HUA2)*. Interaction was also tested using the X-gal filter assay (Figure 3.4B), but as observed in the previous X-gal assay, all assays (except for that testing the CT-HUA2:PRP40(WW) interaction) showed no *LacZ* activation. These results are summarized in Table 3.4, and indicate that both the proline/serine rich region and the proline rich region with the PPLP repeats at the C-terminus are necessary and sufficient for HUA2 to interact with PRP40, RBP45 or UBP1 in yeast cells.

3.3. HUA2 protein family members, HULK1 and HULK3, also interact with HUA2-interacting proteins in yeast cells

The *Arabidopsis* genome contains three genes with high sequence similarity to *HUA2*. These genes are referred to as *HULK* (*HUA2 LIKE*) genes and the family of four genes is referred to as the *HULK* gene family. All four gene family members show similar broad expression domains at the tissue level (Challa, 2009). The *HULK* genes encode proteins whose domain architecture is similar to that of HUA2. Like HUA2, all HULK proteins contain a PWWP domain, an RPR domain, nuclear localization signals, and a proline-rich region. In addition there is one PPLP motif in HULK1 and HULK2, while no such repeats are present in HULK3. The high amino acid sequence similarity, and shared expression patterns among the *HULK* gene family members at the tissue level, indicate possible functional redundancy and suggest that the HULK proteins might also interact with the same splicing factors as HUA2. Also, I was interested to determine whether differences in the number of PPLP motifs affect HULK protein interaction. To investigate this I carried out yeast two-hybrid assays. While HUA2 could be suppressed with 3AT, the HULKs on the other hand could not be supressed, making them unsuitable

Table 3.4. Summary and interpretation of two-hybrid assays conducted with the indicated bait and prey protein.

Fusion gene with				
*GAL*DB	*GAL*AD	*HIS*3 Expression	β-galactosidase Expression	Interpretation
	PRP40(WW)	-	White	No interaction
hua2-1	*RBP45*	-	White	No interaction
hua2-5	*UBP1*	-	White	No interaction
	Empty (N)	-	White	No interaction
CT-HUA2	*PRP40(WW)*(P)	+	Dark Blue	Strong interaction
	RBP45	+	white	Weak interaction
	UBP1	+	White	Weak interaction
	Empty (N)	-	White	No interaction

for yeast two-hybrid analysis (Challa and Pher, unpublished). Since the proline-rich C-terminal region of HUA2 was sufficient for interaction; I decided to examine only the proline-rich regions of the HULKs by yeast two-hybrid analysis. Constructs encoding GAL4 DB-CT-HULKs were prepared in the yeast two-hybrid vector pDBLeu. Next, the self-activation of the fusion proteins was analyzed to determine the concentration of 3AT required to inhibit the basal levels of *HIS*3 reporter gene expression (Figure 3.5). A concentration of 40 mM 3AT was sufficient for inhibiting self-activation of GAL4 DB-CT-HULK1 and GAL4 DB-CT-HULK3 but for GAL4 DB-CT-HULK2 even 100 mM 3AT was not sufficient to inhibit the self-activation, thereby making it unsuitable for our yeast two-hybrid assays. For this reason, yeast two-hybrid analysis was only performed for CT-HULK1 and CT-HULK3. The 3AT concentration required for inhibiting the self-activation of GAL4 DB-CT-HULK1, GAL4 DB-CT-HULK2 and GAL4 DB-CT-HULK3 is listed Table 3.5. Next, the yeast strain MVa203 was co-transformed with plasmids encoding GAL4 DB-CT-HULK1 or GAL4 DB-CT-HULK3 together with plasmids encoding the AD-fusion constructs (WW-PRP40,WW-FCA,UBP1 and RBP45) and assayed for activation of *HIS3* and *Lac*Z transcription. Figure 3.6A shows both CT-HULK1 and CT-HULK3 activated the *HIS*3 reporter gene under stringent conditions (40mM 3AT), indicating their physical interaction. For the CT-HULK1 and UBP1 interaction the colony growth was not as prominent compared to others suggesting only a weak interaction. Although we observed growth of colonies on plates lacking HIS, the X-gal filter assay (Figure 3.6B) showed positive interaction only for CT-HULK1 and PRP40 (WW), CT-HULK1 and RBP45, CT-HULK3 and UBP1 and CT-HULK3 and FCA interactions.

Figure 3.5. Determination of the optimal concentration of 3AT for assaying for GAL4 DB:CT-HULKs fusions. The concentration of 3AT at which the expression of the *HIS3* reporter gene was fully supressed was determined by transforming MVa203 cells expressing the indicated *GAL4 DB*:*CT-HULKs* fusions and plating them on plates lacking leucine and histidine amino acid and containing the indicated concentrations of 3AT.

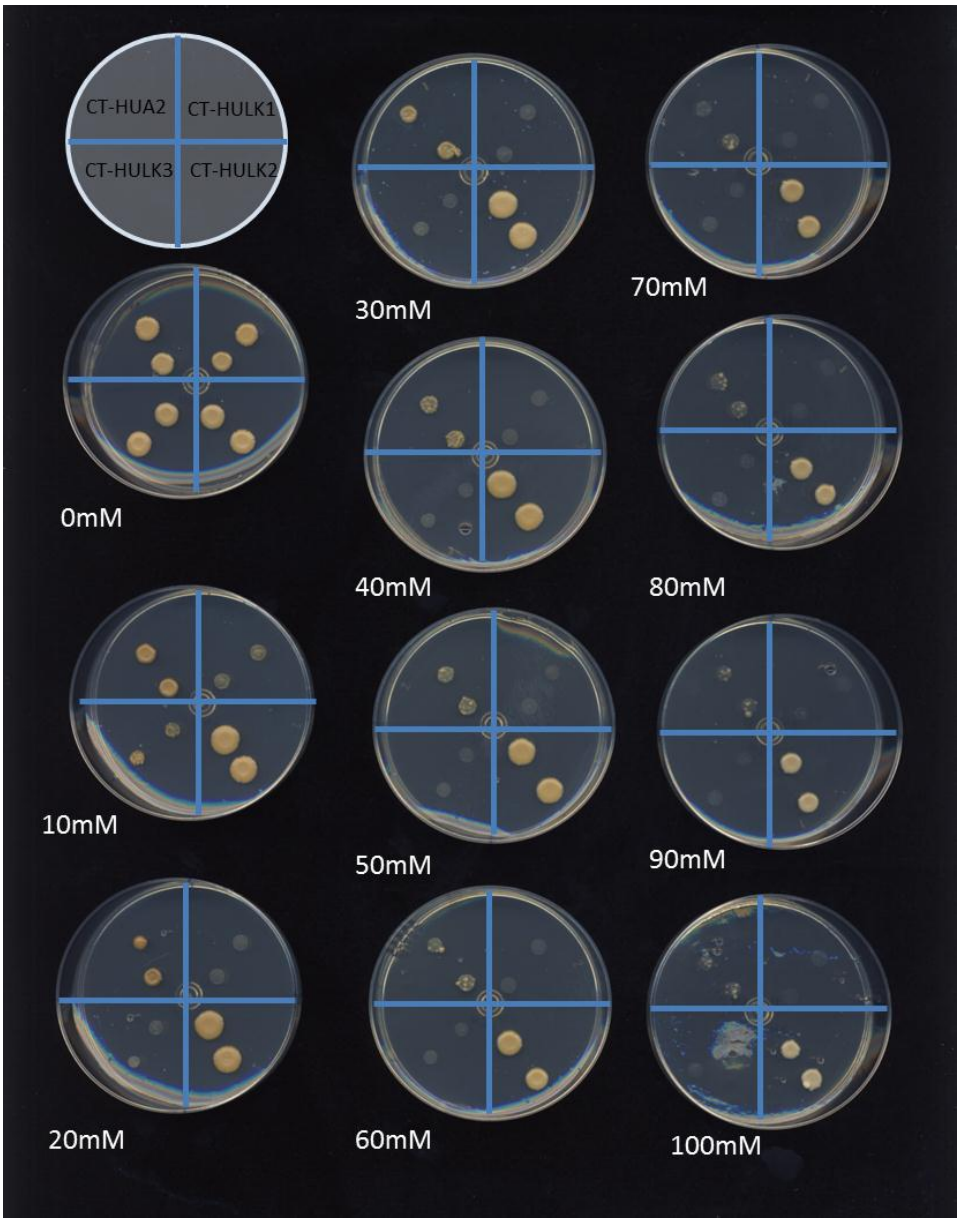
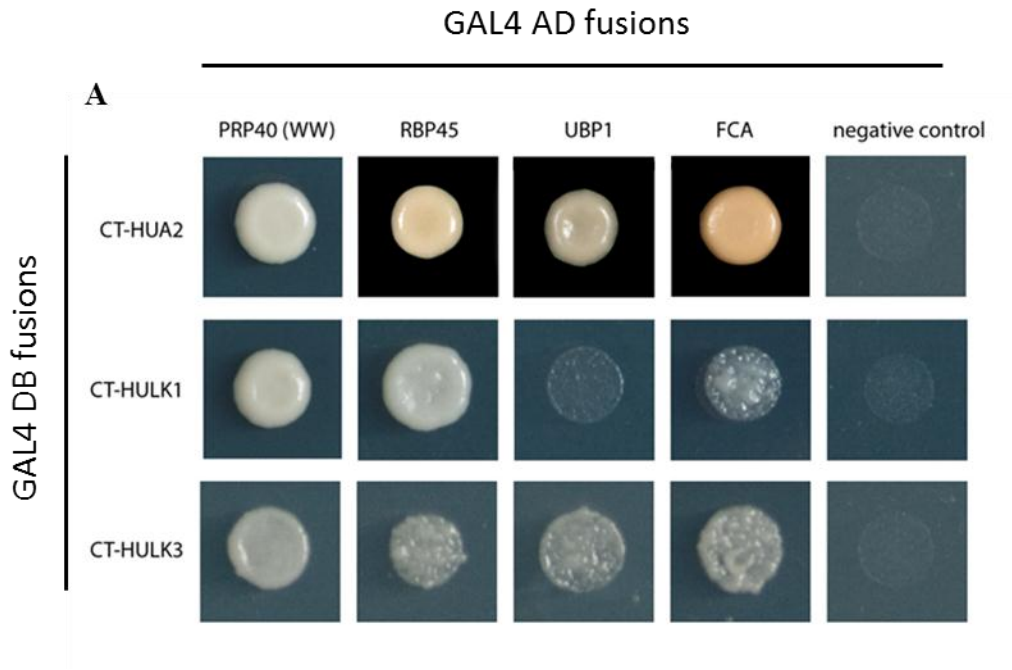

CT-HUA2
CT-HULK1
CT-HULK3
CT-HULK2
0mM
30mM
70mM
40mM
80mM
10mM
50mM
90mM
20mM
60mM
100mM

Table 3.5. 3-Amino-1, 2, 4-Triazol (3AT) concentrations required to inhibit self-activation of GAL4 DB-CT-HUA2 and HULKs fusion proteins.

Fusion gene with *GAL* DB	3 AT concentration required to inhibit self-activation
CT-HUA2	80
CT-HULK1	40
CT-HULK2	<100
CT-HULK3	40

Figure 3.6. Yeast two-hybrid analysis of the interaction of CT-HULK1 and CT-HULK3 with PRP40(WW), RBP45, UBP1 and FCA. MVa203 cells were co-transformed with plasmids expressing the indicated GAL4 DNA-binding domain and GAL4- activation domain fusions **(A)** Colony growth of yeast cells expressing the indicated fusion proteins on plates with 40 mM 3-AT and lacking leucine, tryptophan and histidine (-LTH). (**B**) X-gal filter assays for the colonies shown in (**A**).


GAL4 AD fusions
A
PRP40 (WW)
RBP45
UBP1
FCA
negative control
GAL4 DB fusions
CT-HUA2
CT-HULK1
CT-HULK3

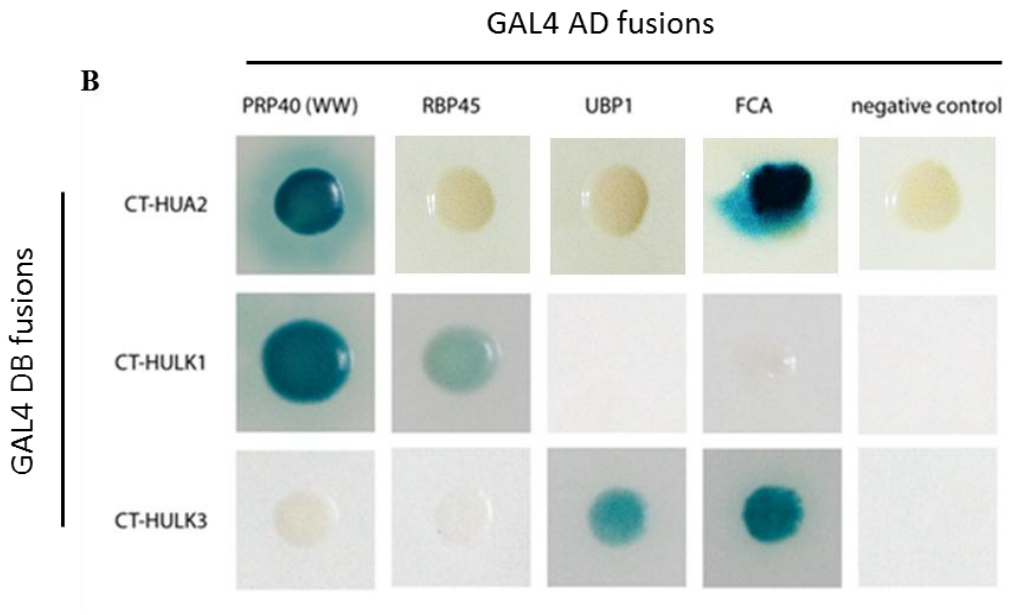

GAL4 AD fusions
B
PRP40 (WW)
RBP45
UBP1
FCA
negative control
GAL4 DB fusions
CT-HUA2
CT-HULK1
CT-HULK3

Table 3.6. Summary and interpretation of yeast two-hybrid assays conducted with the indicated bait and prey proteins.

Fusion gene with *GAL* DB	*GAL*AD	*HIS*3 Expression	β-galactosidase Expression	Interpretation
	PRP40(WW)(P)	+	Blue	Strong interaction
CT-HULK1	*RBP45*	+	Blue	Strong interaction
	UBP1	+	White	Weak interaction
	FCA	+	White	Weak interaction
	Empty (N)	-	White	No interaction
CT-HULK3	*PRP40(WW)*(P)	+	White	Weak interaction
	RBP45	+	White	Weak interaction
	UBP1	+	Blue	Strong interaction
	FCA	+	Blue	Strong interaction
	Empty (N)	-	White	No interaction

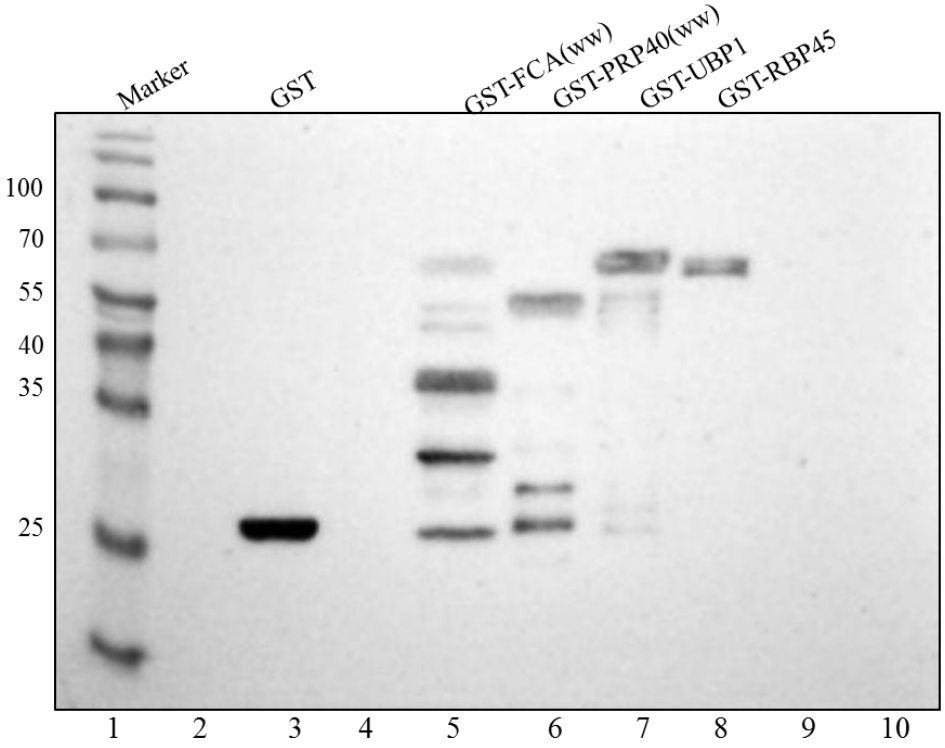
In summary, like HUA2, HULK1 and HULK3 interact with FCA, PRP40, RBP45 and UBP1 in yeast cells. Table 3.6 provides a summary of CT-HULK1 and CT-HULK3 interactions.

3.4 *In vitro* interaction between the HULK proteins and PRP40, RBP45 and UBP1

In order to confirm the data obtained through two-hybrid analysis, I performed *in vitro* glutathione *S*-transferase (GST) pull-down assays. This assay analyzes the binding of HIS-tagged prey proteins (HUA2, HULK1, HULK2 or HULK3) to a purified GST-tagged bait protein (FCA, PRP40, RBP45 or UBP1) bound to glutathione-Sepharose beads. My attempt to express full length HUA2 protein was not successful. Since the C-terminal regions of HUA2, HULK1 and HULK3 proteins were sufficient for interacting with FCA, PRP40, RBP45 or UBP1 in yeast cells I decided to use only the C-terminal portions of the HUA2, HULK1 and HULK3 proteins as prey for the GST pull-down assays. Since CT-HULK2 was unsuitable for analysis in the yeast two-hybrid assay, I decided to include it as part of my GST pull-down assay. Also, only the WW domains of PRP40 and FCA (which is sufficient for interaction with the HUA2 protein family in yeast cells) were used as bait proteins.

Fusion proteins, GST-WW-PRP40, GST-RBP45, GST-UBP1 and GST-WW-FCA as well as GST alone (negative control) were expressed in *E. coli*. Expression of the fusions was examined by immunoblotting with anti-GST antibodies (Figure 3.7). High levels of expression were obtained for all fusion proteins except the WW domain of FCA. Several non-specific bands were visible for GST-WW-FCA. I tried expressing a full length GST-FCA fusion but it also did not express in *E. coli* (data not shown).

Figure 3.7. Western blot analysis to determine the expression of GST and GST tagged proteins. Protein marker (lane 1), GST (lane 3), GST-FCA (WW) (lane 5), GST-PRP40 (WW) (lane 6), GST-UBP1 (lane 7), and GST-RBP45 (lane 8). Bacterially expressed GST fusion proteins were analyzed by SDS-PAGE followed by immunoblotting with anti-GST antibodies.


Marker
GST
GST-FCA(ww)
GST-PRP40(ww)
GST-UBP1
GST-RBP45
100
70
55
40
35
25
1
2
3
4
5
6
7
8
9
10

Anti GST Ab

Hence, I was unable to use FCA for the GST pull-down assays. Next, the expression of HIS tagged prey proteins in *E. coli* was determined by immunoblotting with anti-HIS antibodies (Figure 3.8). Immunoblotting was also used to confirm that no non-specific binding was taking place between anti GST antibodies and HIS fusion proteins or between anti-HIS antibodies and GST fusion protein (Figure 3.9 and 3.10).

Next I performed GST pull-down control experiments to determine if the HIS tagged prey proteins, HIS-CT-HUA2, HIS-CT-HULK1, HIS-CT-HULK2 or HIS-CT-HULK3 bind to GST alone or to the glutathione-Sepharose 4B beads. As seen in Figure 3.11-3.12 I confirmed that none of the HIS fusion prey proteins bind to GST or to the glutathione-Sepharose 4B. After confirming that there was no non-specific binding of the prey proteins to GST, I first examined the binding of CT-HUA2 with WW-PRP40, RBP45 and UBP1**.** Figure 3.13 shows the retention of HIS-CT-HUA2 by GST-WW-PRP40, GST-RBP45 and GST-UBP1 on glutathione-Sepharose beads. The fact that HIS-CT-HUA2 does not bind to GST or to the glutathione-Sepharose beads in our control experiments (Figure 3.11) indicates that the binding between CT-HUA2 and PRP40, RBP45 and UBP1 is indeed specific.

Next, I determined if the CT-HULKs interact with PRP40. Figure 3.14 shows the retention of HIS-CT-HUA2 (positive control), HIS-CT-HULK1, HIS-CT-HULK2, HIS-CT-HULK3 by GST-WW-PRP40 on glutathione-Sepharose beads. The fact that HIS-CT-HULK1, HIS-CT-HULK2, HIS-CT-HULK3 does not bind to GST alone or to the glutathione-Sepharose beads in control experiments shown in Figure 3.11 and 3.12 indicates that the binding between CT-HULK1, CT-HULK2, CT-HULK3 and

Figure 3.8. Western blot analysis to determine the expression of HIS tagged prey proteins. Protein marker (lane 1), HIS-CT-HUA2 (lane 3), HIS-CT-HULK1 (lane 5), HIS-CT-HULK2 (lane 7) and HIS-CT-HULK3 (lane 9). Bacterially expressed HIS fusion proteins were analyzed by SDS-PAGE followed by immunoblotting with anti-HIS antibodies.

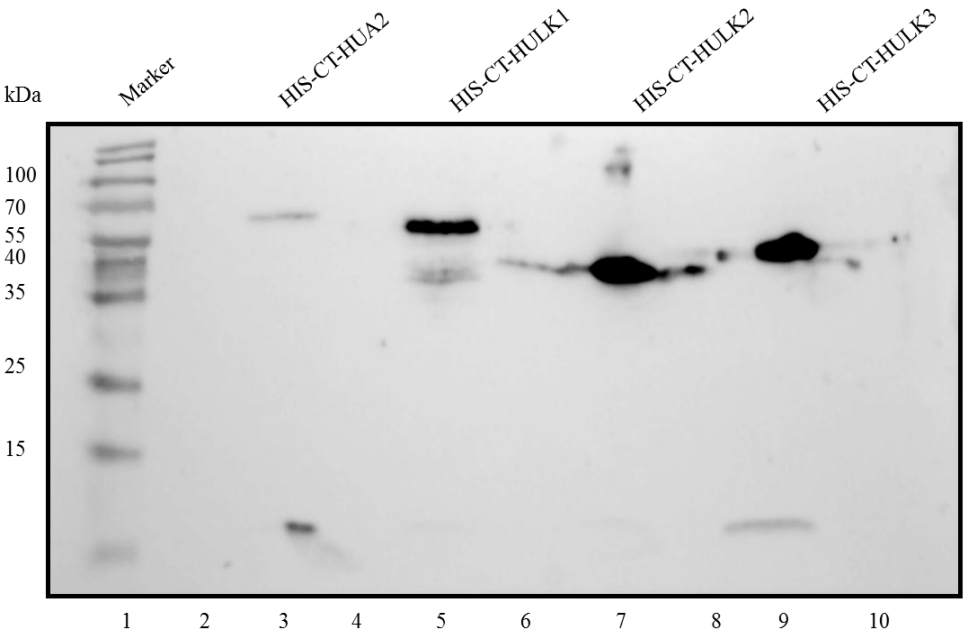

kDa
Marker
HIS-CT-HUA2
HIS-CT-HULK1
HIS-CT-HULK2
HIS-CT-HULK3
100
70
55
40
35
25
15
1
2
3
4
5
6
7
8
9
10
Anti HIS Ab

Figure 3.9. Western blot analysis to determine HIS-CT-HUA2, HIS-CT-HULK1, HIS-CT-HULK2 and HIS-CT-HULK3 cross-reaction with anti-GST antibody. Lysate containing 30 µg of total protein for HIS-CT-HUA2 (lane 2), HIS-CT-HULK1(lane 4), HIS-CT-HULK2 (lane 6), HIS-CT-HULK3 (lane 8), GST alone (lane 9) and protein ladder (lane M) was first resolved by SDS-PAGE and then transferred to a nitrocellulose membrane, followed by immunoblotting with **(A)** anti-HIS antibody. **(B)** The same membrane was stripped and reprobed with the anti-GST antibody to confirm that anti-GST antibody reacts only with the control GST but not with any HIS tagged prey protein.

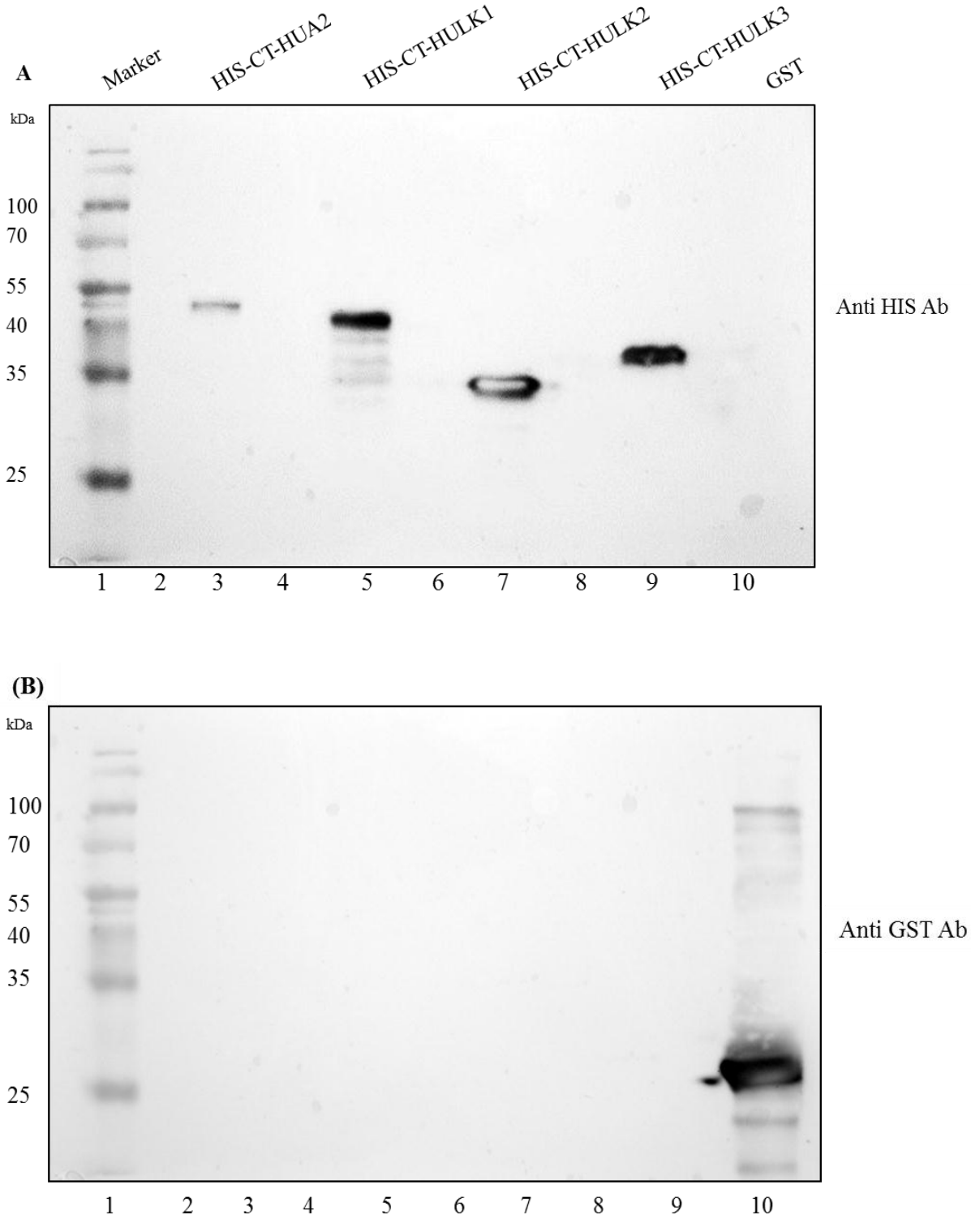

A
Marker
HIS-CT-HUA2
HIS-CT-HULK1
HIS-CT-HULK2
HIS-CT-HULK3
GST
kDa
100
70
55
40
35
25
1 2 3 4 5 6 7 8 9 10
Anti HIS Ab
(B)
kDa
100
70
55
40
35
25
1 2 3 4 5 6 7 8 9 10
Anti GST Ab

Figure 3.10. Western blot analysis to determine GST, GST-AtCTD, GST-PRP40(WW), GST-RBP45 and GST-UBP1 cross reaction with anti-HIS antibody. Lysate containing 30 µg of total protein for GST (lane 2), GST-AtCTD (lane 3), GST-PRP40(WW) (lane 4), GST-RBP45 (lane 5), GST-UBP1 (lane 6), for control HIS-CT-HUA2 (lane 9) and protein ladder (lane M) was first resolved by SDS-PAGE and then transferred to a nitrocellulose membrane, followed by immunoblotting with **(A)** anti-GST antibody **(B)** The same membrane was stripped and reprobed with anti-HIS antibody to confirm that anti-HIS antibody reacts only with the control HIS-CT-HUA2 but not with any GST tagged bait proteins.

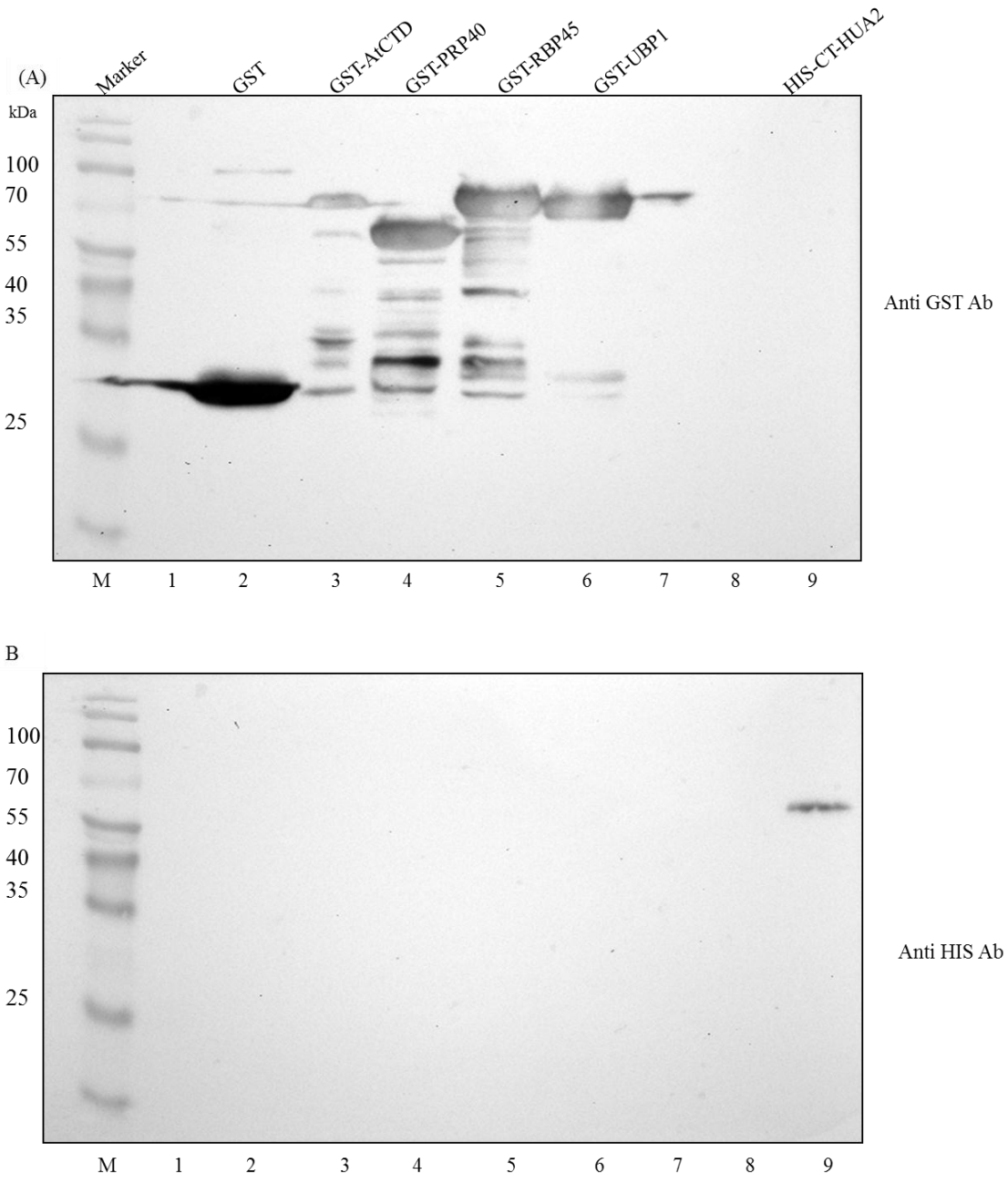

(A)
Marker
GST
GST-AtCTD
GST-PRP40
GST-RBP45
GST-UBP1
HIS-CT-HUA2
kDa
100
70
55
40
35
25
M 1 2 3 4 5 6 7 8 9
Anti GST Ab
B
100
70
55
40
35
25
M 1 2 3 4 5 6 7 8 9
Anti HIS Ab

Figure 3.11. Western blot analysis to determine CT-HUA2 and CT-HULK1 binding to GST or to glutathione-Sepharose 4B beads. (A) Controls for *in vitro* glutathione *S*-transferase (GST) pull-down assays. GST protein was immobilized on glutathione beads and incubated alone (lane 1) or with HIS-CT-HUA2 (lane 3), or CT-HULK1 (lane 7). The same amount of HIS-CT-HUA2 (lane 4) and HIS-CT-HULK1 (lane 8) was incubated with only glutathione Sepharose 4B beads. Lane 5 and lane 9 contains lysate containing HIS-CT-HUA2 and CT-HULK1 respectively. The bound proteins were separated by SDS-PAGE transferred to a nitrocellulose membrane and then immunobloted with anti-GST antibody **(B)** The nitrocellulose membrane described in **(A)** was stripped and reprobed with anti-HIS antibody.

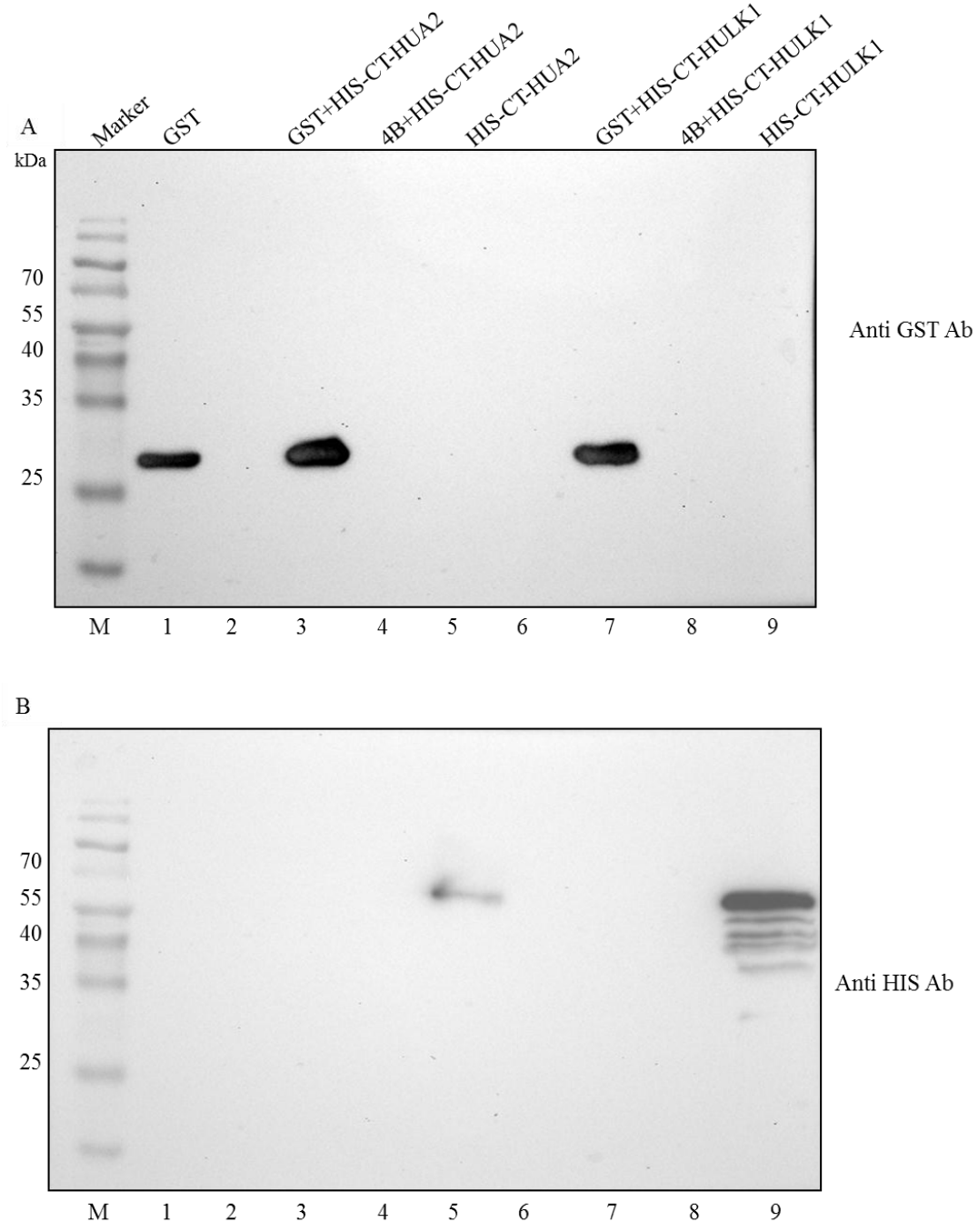

A
Marker
GST
GST+HIS-CT-HUA2
4B+HIS-CT-HUA2
HIS-CT-HUA2
GST+HIS-CT-HULK1
4B+HIS-CT-HULK1
HIS-CT-HULK1
kDa
70
55
40
35
25
M 1 2 3 4 5 6 7 8 9
Anti GST Ab
B
70
55
40
35
25
M 1 2 3 4 5 6 7 8 9
Anti HIS Ab

Figure 3.12. Western blot analysis to determine CT-HULK2 and CT-HULK3 binding to GST or glutathione- Sepharose 4B beads. (**A**) Controls for *in vitro* glutathione *S*-transferase (GST) pull-down assays. GST protein was immobilized on glutathione beads and incubated alone (lane 1) or with lysate containing HIS-CT-HULK2 (lane 3), or CT-HULK3 (lane 7). The same amount of HIS-CT-HULK2 (lane 4) and HIS-CT-HULK3 (lane 8) was incubated with only glutathione-Sepharose 4B beads. Lane 5 and lane 9 contains lysate containing HIS-CT-HULK2 and CT-HULK3 respectively The bound proteins were separated by SDS-PAGE, transferred to a nitrocellulose membrane and then immunoblotted with anti-GST antibody **(B)** the same membrane was stripped and reprobed with anti-HIS antibody.

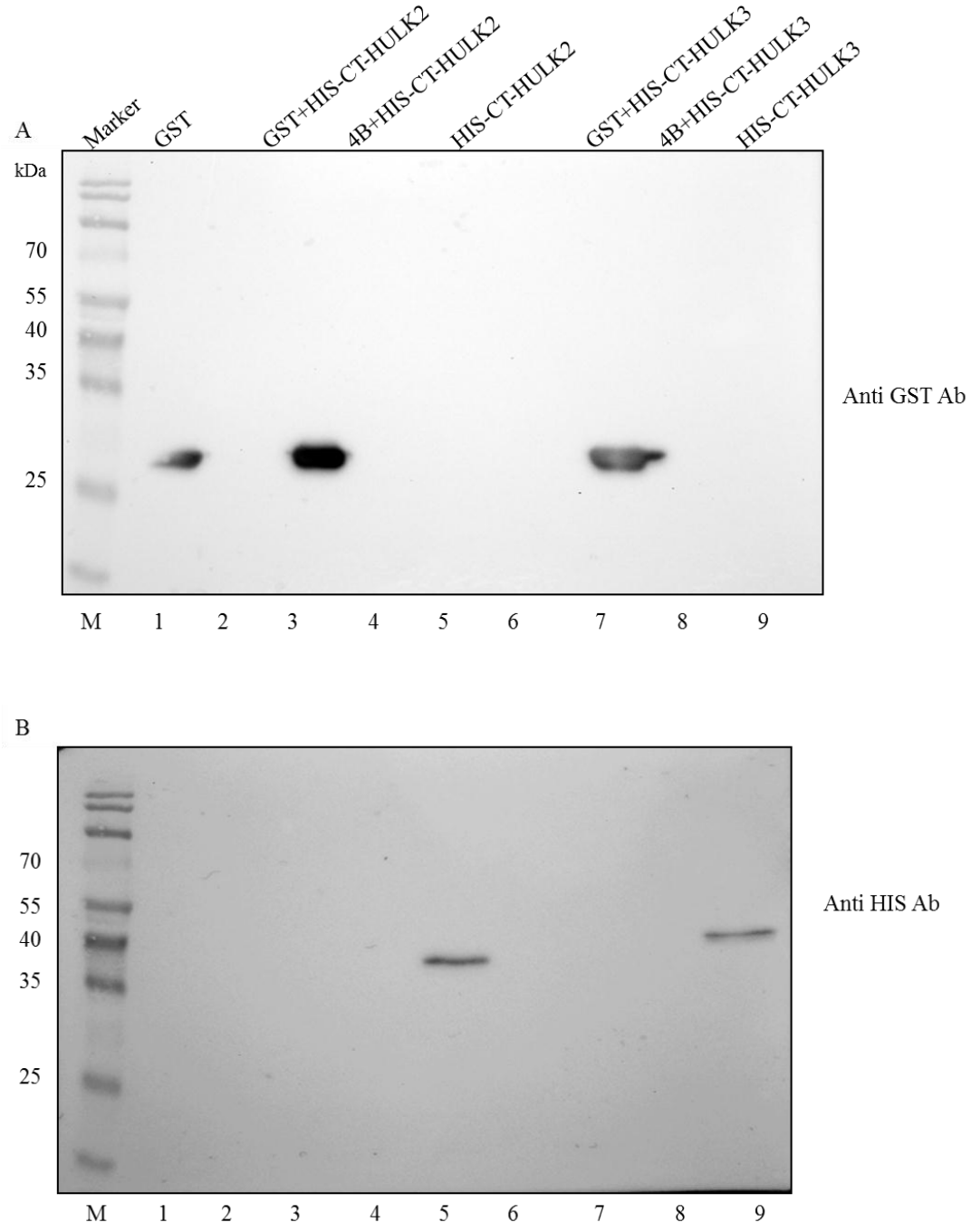

A
Marker
GST
GST+HIS-CT-HULK2
4B+HIS-CT-HULK2
HIS-CT-HULK2
GST+HIS-CT-HULK3
4B+HIS-CT-HULK3
HIS-CT-HULK3
kDa
70
55
40
35
25
Anti GST Ab
M 1 2 3 4 5 6 7 8 9
B
70
55
40
35
25
Anti HIS Ab
M 1 2 3 4 5 6 7 8 9

Figure 3.13. *In vitro* Analysis of HUA2 interaction with PRP40, RBP45 and UBP1 by GST pull-down assays. Either GST-PRP40(WW) (lane 2 and 6), GST-RBP45 (lane 3 and 7) or GST-UBP1 (lane 4 and 8) bound to glutathione-Sepharose 4B was incubated with HIS-CT-HUA2. Lane 9 was loaded with lysate containing HIS-CT-HUA2 as a reference marker. Bound proteins were eluted and separated by SDS-PAGE stained with coomassie **(A)** then transferred to a nitrocellulose membrane, followed by immunoblotting with anti-GST antibody **(B)** and with anti-HIS antibody **(C)**.

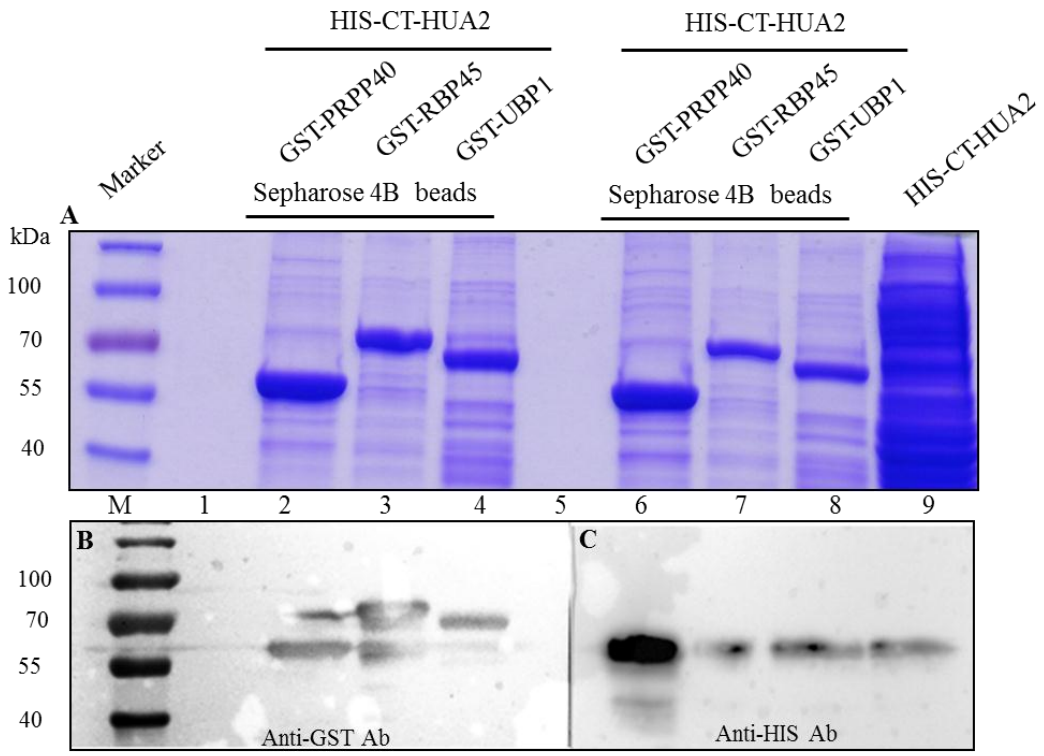

HIS-CT-HUA2
HIS-CT-HUA2
GST-PRPP40
GST-RBP45
GST-UBP1
GST-PRPP40
GST-RBP45
GST-UBP1
Marker
HIS-CT-HUA2
Sepharose 4B beads
Sepharose 4B beads
A
kDa
100
70
55
40
M 1 2 3 4 5 6 7 8 9
B
C
100
70
55
40
Anti-GST Ab
Anti-HIS Ab

Figure 3.14. *In vitro* analysis of CT-HULK1, CT-HULK2 and CT-HULK3 interaction with PRP40(WW) by GST pull-down assays GST-PRP40(WW) bound to glutathione-Sepharose 4B was incubated with HIS-CT-HUA2 (lane 2), HIS-CT-HULK (lane 4), HIS-CT-HULK2 (lane 6) or with HIS-CT-HULK3 (lane 8). For reference, lysate containing HIS-CT-HUA2 (lane 3), HIS-CT-HULK1 (lane 5), HIS-CT-HULK2 (lane 7) and HIS-CT-HULK3 (lane 9) was loaded. Bound proteins were eluted and separated by SDS-PAGE stained with coomassie **(A)** then transferred to a nitrocellulose membrane, followed by immunoblotting with anti-GST antibody **(B)** or with anti-HIS antibody **(C)**.

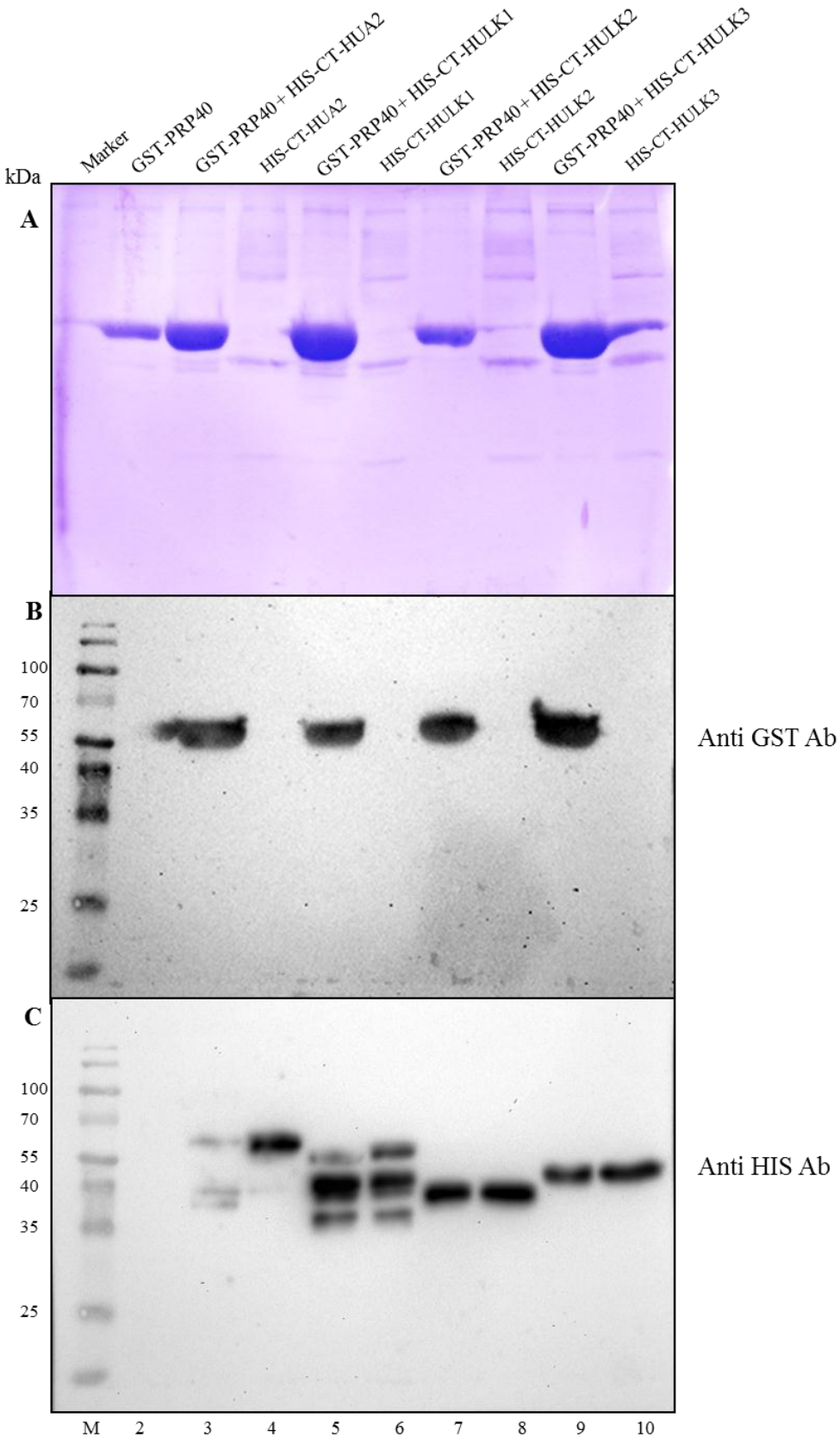
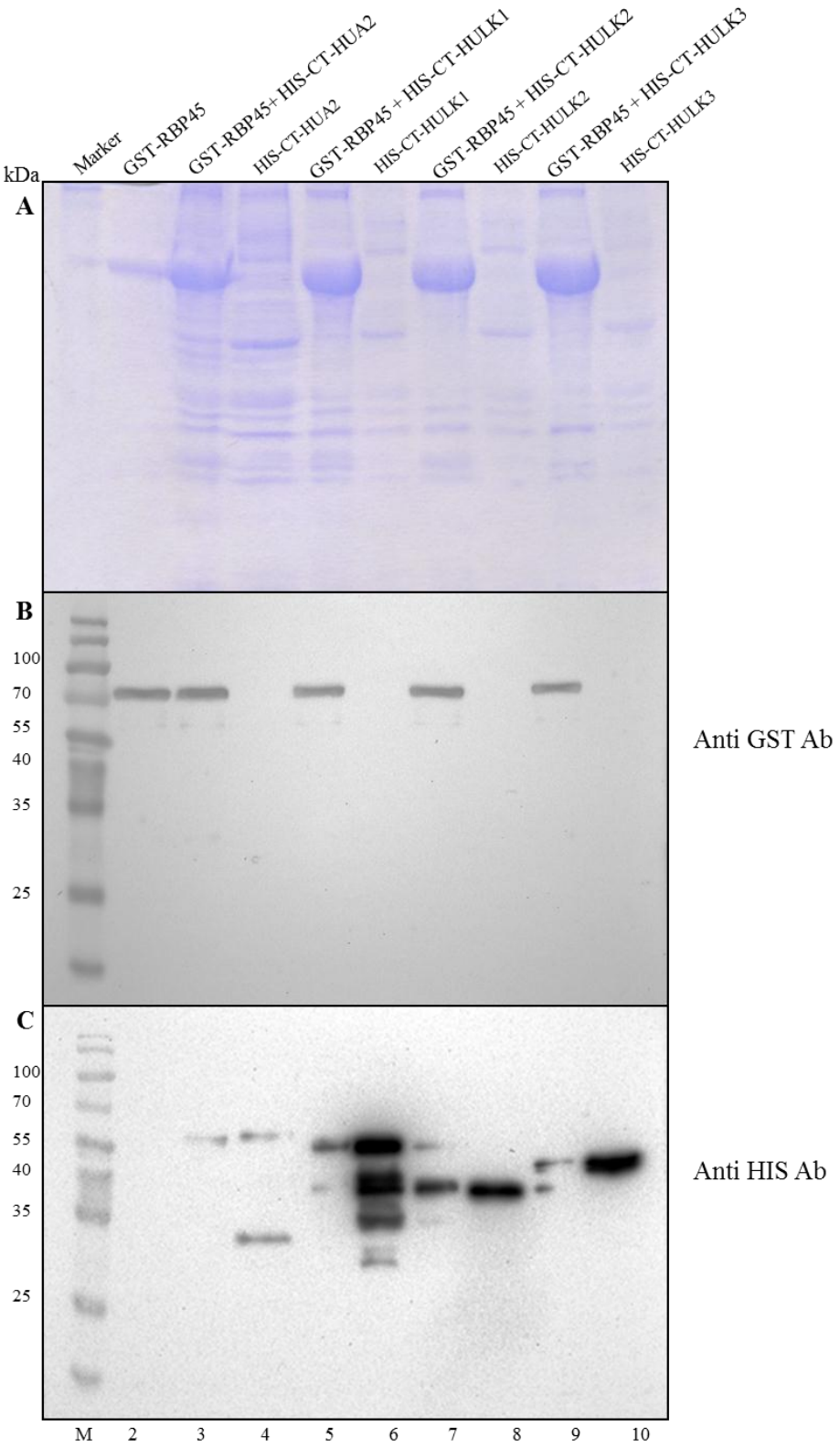

Marker
GST-PRP40
GST-PRP40+ HIS-CT-HUA2
HIS-CT-HUA2
GST-PRP40+ HIS-CT-HULK1
HIS-CT-HULK1
GST-PRP40+ HIS-CT-HULK2
HIS-CT-HULK2
GST-PRP40+ HIS-CT-HULK3
HIS-CT-HULK3
kDa
A
B
100
70
55
40
35
25
Anti GST Ab
C
100
70
55
40
35
25
Anti HIS Ab
M 2 3 4 5 6 7 8 9 10

Figure 3.15. ***In vitro*** **analysis of CT-HULK1, CT-HULK2 and CT-HULK3 interaction with RBP45 by GST pull-down assays**. GST-RBP45 bound to glutathione-Sepharose 4B was incubated with HIS-CT-HUA2 (lane 2), HIS-CT-HULK (lane 4), HIS-CT-HULK2 (lane 6), or with HIS-CT-HULK3 (lane 8). For reference, lysate containing HIS-CT-HUA2 (lane 3), HIS-CT-HULK1 (lane 5), HIS-CT-HULK2 (lane 7) and HIS-CT-HULK3 (lane 9) was loaded. Bound proteins were eluted and separated by SDS-PAGE stained with coomassie **(A)** then transferred to a nitrocellulose membrane, followed by immunoblotting with anti-GST antibody **(B)** or with anti-HIS antibody **(C)**.


Marker
GST-RBP45
GST-RBP45+ HIS-CT-HUA2
HIS-CT-HUA2
GST-RBP45 + HIS-CT-HULK1
HIS-CT-HULK1
GST-RBP45 + HIS-CT-HULK2
HIS-CT-HULK2
GST-RBP45 + HIS-CT-HULK3
HIS-CT-HULK3
kDa
A
B
100
70
55
40
35
25
Anti GST Ab
C
100
70
55
40
35
25
Anti HIS Ab
M 2 3 4 5 6 7 8 9 10

PRP40, are indeed specific. Similar pull-down assays were performed to confirm the interaction between RBP45 and CT-HULKs (Figure 3.15) and between UBP1 and CT-HULKs (Figure 3.16). In this last pull-down experiment I was unable to demonstrate an interaction between GST-UBP and HIS-CT-HUA2. Taken together, our data indicate that the HULK protein family directly interacts both *in vivo* and *in vitro* with PRP40, RBP45 and UBP1.

3.5 HUA2 does not form homodimer or heterodimers

Our laboratory has previously demonstrated a role for *HUA2* and *HULKs* in splicing. The fact that this protein family interacts with splicing factors *in vivo* and *in vitro* suggests that the HULK protein family members might be participating in splicing. Many plant transcription factors are known to form hetero- or homoidmers (Mingyu et al., 2012). Also, Yang et al. (2003) reported that the PPLP domain plays a key role in multimerization. The fact that most of the HULK protein family members have PPLP domains made us question if the members of the HUA2 family form homo- or heterodimers. To determine if this were true I carried out yeast two-hybrid analysis by using both full length *HUA2* and the carboxyl-terminal region of *HUA2* containing the proline rich region with PPLP repeats. I co-transformed MVa203 cells with plasmids expressing the *GAL4* DNA-binding domain fused to *HUA2* or *CT-HUA2* together with plasmids encoding the relevant GAL4 AD fusions and assayed for activation of *HIS3* and *Lac*Z transcription. Figure 3.17 (A1 and B1) shows that both *HUA2* and *CT-HUA2* do not interact in yeast under stringent conditions (50 mM 3AT for *HUA2* and 80 mM 3AT for *CT-HUA2)*. These results were confirmed by an X-gal filter assay (Figure 3.17 A2 and B2) and are summarized in Table 3.7. In summary, these results suggest that HUA2

Figure 3.16. ***In vitro*** **analysis of CT-HULK1, CT-HULK2 and CT-HULK3 interaction with UBP1 by GST pull-down assays**. GST-UBP1 bound to glutathione-Sepharose 4B was incubated with HIS-CT-HUA2 (lane 2), HIS-CT-HULK (lane 4), HIS-CT-HULK2 (lane 6) or with HIS-CT-HULK3 (lane 8). For reference lysate containing HIS-CT-HUA2 (lane 3), HIS-CT-HULK1 (lane 5), HIS-CT-HULK2 (lane 7) and HIS-CT-HULK3 (lane 9) was loaded. Bound proteins were eluted and separated by SDS-PAGE stained with coomassie **(A)** then transferred to a nitrocellulose membrane, followed by immunoblotting with anti-GST antibody **(B)** or with anti-HIS antibody **(C)**.

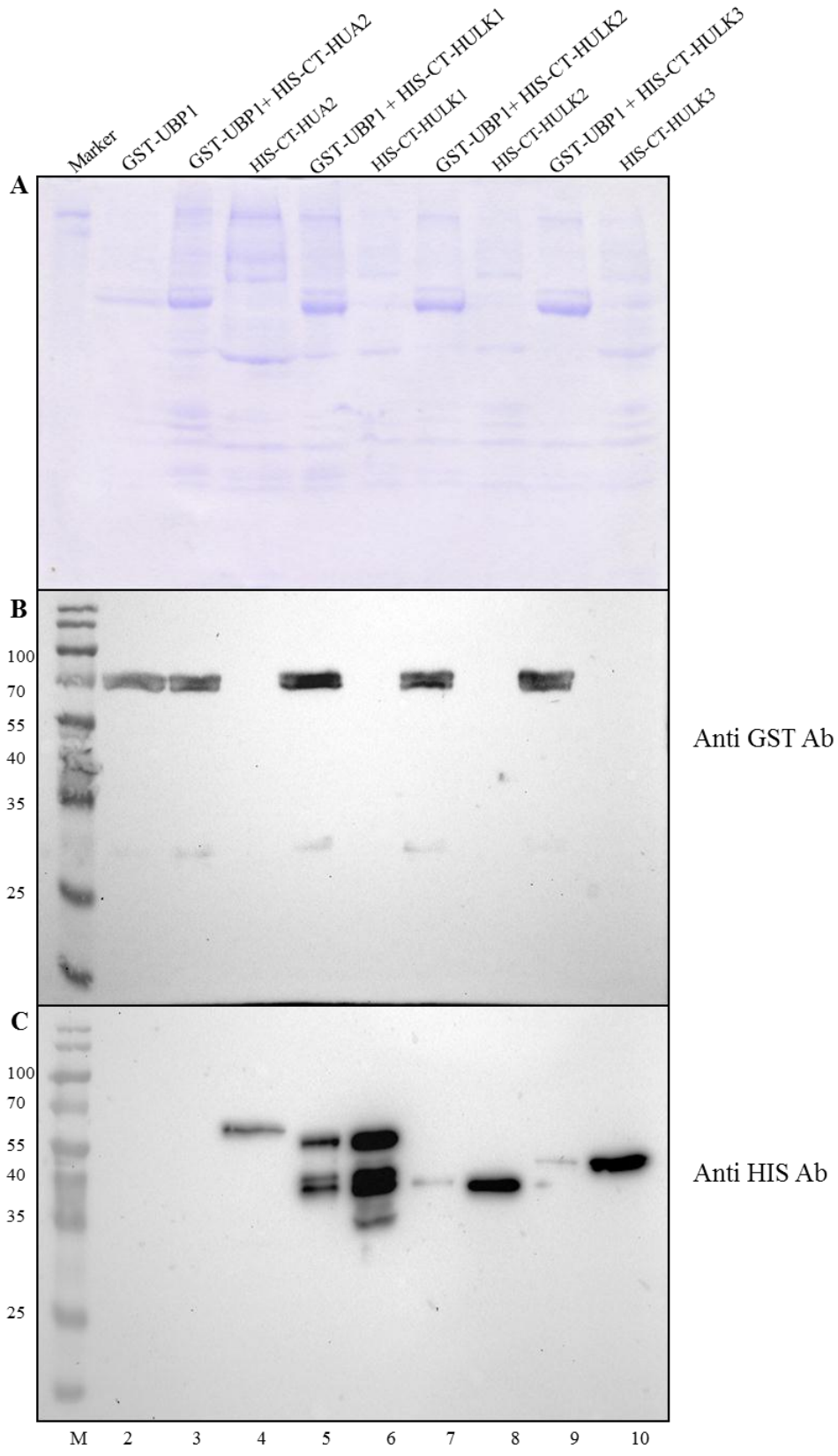

Marker
GST-UBP1
GST-UBP1+ HIS-CT-HUA2
HIS-CT-HUA2
GST-UBP1 + HIS-CT-HULK1
HIS-CT-HULK1
GST-UBP1+ HIS-CT-HULK2
HIS-CT-HULK2
GST-UBP1 + HIS-CT-HULK3
HIS-CT-HULK3
A
B
100
70
55
40
35
25
Anti GST Ab
C
100
70
55
40
35
25
Anti HIS Ab
M 2 3 4 5 6 7 8 9 10

Figure 3.17. Yeast two-hybrid analysis of the interaction of HUA2 and CT-HUA2 with HUA2, HULK1, HULK2 or HULK3. MVa203 cells were co-transformed with constructs encoding GAL4 DB-HUA2 **(A1)** and GAL4 DB-CT-HUA2 **(B1)** along with constructs encoding GAL4-AD:HUA2, GAL4-AD-HULK1, GAL4-AD:HULK2 and GAL4-AD-HULK3. Transformants were grown on selection plates lacking leucine, tryptophan and histidine (-LTH) with 50 mM 3-AT for GAL4 DB:HUA2 and 80 mM 3-AT for GAL4 DB-CT-HUA2. X-gal filter assay for GAL4 DB-HUA2 and GAL4 DB-CT-HUA2 are shown in panels **(A2) (B2)**.

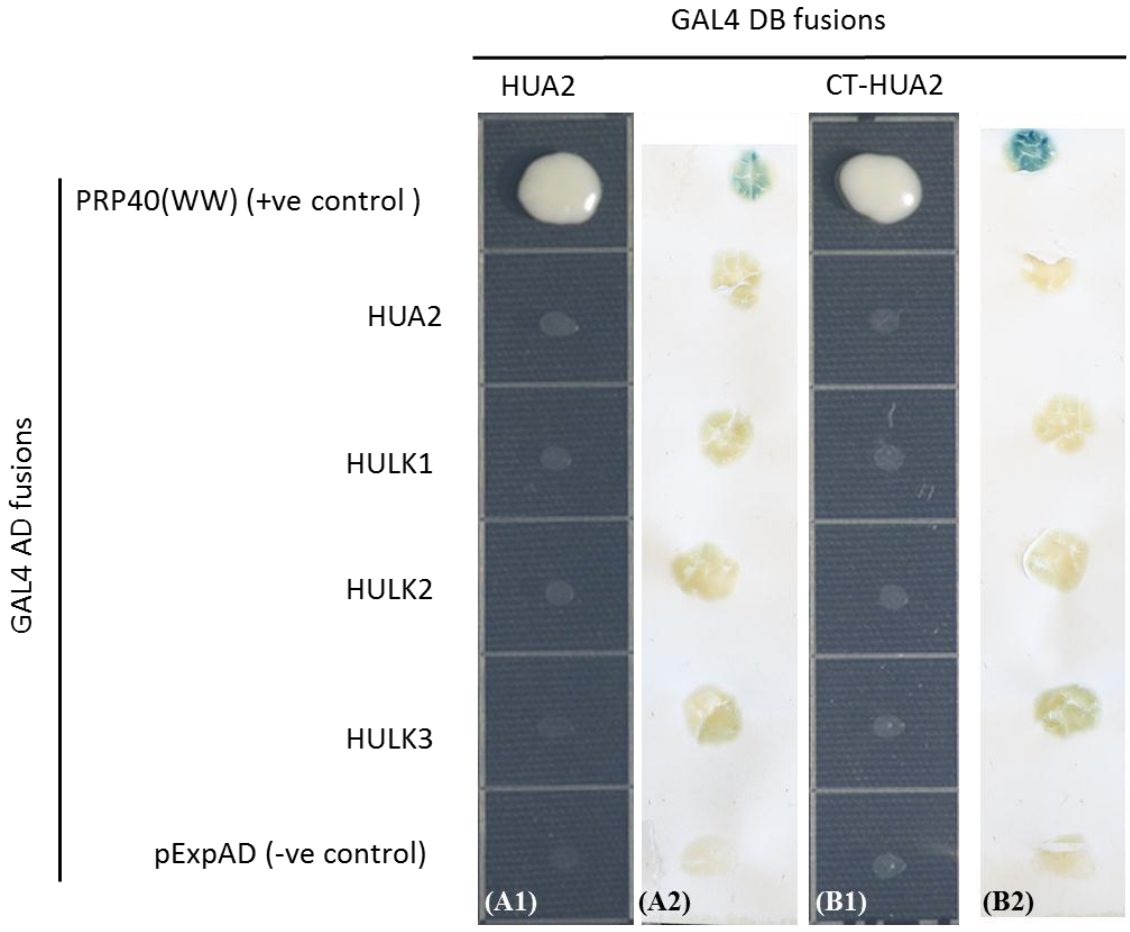

GAL4 DB fusions
HUA2
CT-HUA2
GAL4 AD fusions
PRP40(WW) (+ve control)
HUA2
HULK1
HULK2
HULK3
pExpAD (-ve control)
(A1)
(A2)
(B1)
(B2)

Table 3.7. Summary and interpretation of yeast two-hybrid assays conducted with the indicated bait and prey proteins

Fusion gene with				
*GAL*DB	*GAL*AD	*HIS*3 Expression	β-galactosidase Expression	Interpretation
	PRP40(WW)(P)	+	Blue	Strong interaction
HUA2	*HUA2*	-	White	No interaction
	HULK1	-	White	No interaction
	HULK2	-	White	No interaction
	HULK3	-	White	No interaction
	Empty (N)	-	White	No interaction
CT-HUA2	*PRP40*(WW) (P)	+	Blue	Strong interaction
	HUA2	-	White	No interaction
	HULK1	-	White	No interaction
	HULK2	-	White	No interaction
	HULK3	-	White	No interaction
	Empty (N)	-	White	No interaction

does not self-associate to form a homodimer nor does it interact with any of its protein family members to form a heterodimer in yeast cells.

3.6 Members of HUA2 protein family localize within the nucleus

The presence of three domains rich in acidic residue at the C terminus of HUA2 suggests that HUA2 may be a transcription factor (Chen and Meyerowitz, 1999). Data from our laboratory suggests that HUA2 and the HULKs are involved in splicing of *AG* and *FCA* (Wang, Sajja, Challa and Grbic, unpublished results). Lastly, I have shown that HUA2 and the HULKs interact with splicing factors, PRP40, RBP45, and UBP1 both *in vivo* and *in vitro*. If HUA2 and the HULKs have a direct role to play in the regulation of transcription, it follows that they should be localized to the nucleus. I first carried out an *in silico* analysis using the program NucPred (Brameier et al., 2007) to determine the localization prediction for HUA2 and the HULKs. NucPred predicted that members of the HUA2 protein family localize to the nucleus (Figure 3.18). To determine the subcellular localization of HUA2 and HULK proteins in plant cells, I performed *in vivo* targeting experiments in tobacco (*Nicotiana benthamiana*) leaf tissue. To this end, 35S:HUA2-GFP, 35S:HULK1-YFP, 35S:HULK2-YFP and 35S:HULK3-YFP constructs were generated and delivered into leaf cells of tobacco by *Agrobacterium tumefaciens* infiltration (de Felippes and Weigel, 2010). Co-expression of plasmids encoding mCherry-NLS served as a positive control for localization while co-expression of a plasmid encoding p19 prevented post-transcriptional gene silencing (PTGS) (Karimi et al., 2002; Shaner et al., 2004; Voinnet et al., 2000; Harter, unpublished). In case of HUA2, a strong green fluorescent protein (GFP) signal was observed within the nucleus

Figure 3.18. NucPred coloured multiple sequence alignment for HUA2 protein family members. The consensus ('cons') is calculated as follows: the most frequent character in each column is shown in lowercase (unless it is a gap character); uppercase letters represent a column containing just one amino acid and no gaps. Positively and negatively influencing subsequences are coloured according to the following scale:

(non-nuclear) negative  positive (nuclear)

Sequences with a score >= to a threshold of 0.8 are predicted to spend more time in the nucleus.

Note: the scores for the genes are next to their respective sequence in the Figure.

```
HUA2     0.99 MAPGRKRGASKAKAKG----QLVLGDLVLAKVKGFPAWPAKISRPEDWDRAPDPKKYFVQ
HULK1    0.98 MAPGRKRGANKAMAIG----EMRLGDLVLAKVKGFPAWPAKIGQPEDWNQAPDPKKHFVQ
HULK2    1.00 MAPSRRKGGGRAAAVAAACRKRKVGDLVLAKVKGFPAWPAVVSEPEKWDASPDSKKVFVH
HULK3    0.99 MAPSRKRGGGRAAAASSARREWKVGDLVLAKVKGFPAWPAVVDEPEKWGHSADSKKVTVH
         cons MAP RkrG g AaA g    e k GDLVLAKVKGFPAWPA  sePE Wd  pD KKvfV

HUA2     0.99 FFGTEEIAFVAPPDIQAFTSEAKSKLLARCQGKTVKYFAQAVEQICTAFEGLQNH-KSN-
HULK1    0.98 FYGTGEIGFVTPPDIQPFTSETKKKLSARCQGKTVKYFSQAVEEISAAFEESQKQ-KSD-
HULK2    1.00 FFGTQQIAFCNPGDVEAFTEERKQSLLTRRHAKGSD-FVRAVKEIIESYEKLKQQERASD
HULK3    0.99 FFGTQQIAFCNHGDVESFTEEKKQSLLTRRHAKGSD-FVRAVKEITESYEKLKQQDQASG
         cons FfGTq IaF np D  aFT E Kq Ll R   K    Fv AV eI e  Ekl qq k s

HUA2     0.99 -ALGDEDSLDATEPGLTKAEIVDGTDHIVIESERTDNFNFRVDPCFPKLDENNGEERKAE
HULK1    0.98 -IVGNEALLNAVEPSVTKPKYLN----QASSDGKSDKFSSRADPCLGKLVENNGAEINPD
HULK2    1.00 PKSAEEGTLGSAENTTLMPQVIEIPTATSL-TQMNSDPSHGRDESTLLNEDASAAEQMLA
HULK3    0.99 PKYAEETTAGSSGNTSQLPQACENLIGSRLDTQIESSSSHGRDELTLLSEDASAAEQMLA
         cons  k  eE tlg  e t tkpq  e      l tq    fsh rD ctl le    aEqmla

HUA2     0.99 IRKLDSSSFLESKVKTTSPVSESLEHSSFDPKIKKEDFDKGTDGSACNEHFGNGQKKLAN
HULK1    0.98 VGEQDSS---ISNNRNTSPSSEPVEHGSPDP-ILKVAVDDKIDNVTCTDHSDGTGNNLVN
HULK2    1.00 LRDNSGPRNKACDSAVVKEPRKIATYSSRKRNGGVRSQNCAPQNETCPVQRSKSPSRLQT
HULK3    0.99 LRHNTLAHNGACDSAAAKDLCEIATYSSRRRNERVRALKYAPQSIILPVEHSKISSRLEL
         cons lr ndss n a dsa t p seia  sSrd ni  ra d ap n tcpvh sk  srL n

HUA2     0.99 GKR-IKKEAGGSDRKGEDTVHRDKSNNS------HVPG------GRTASGNSDSKKSK--
HULK1    0.98 DQRIIRKTTDDSNKRCKDEVRAKRVPDSRAATDNHILGPNQKLKGSIKGQDHGSKKGQDH
HULK2    1.00 EKLQSSMLQNSDGGQTIDDVEDGALRREK-----RIRR----SSGHSESDDVATSSLNSH
HULK3    0.99 DRVQRSMLQCSDGGPSVNSINGKAIRRRK-----RIR-----TSGQSESDDVVSSDLNLH
         cons dkrqis lq s gg   d v  ka rrsk      irg     sG sesddv s kln h

HUA2     0.99 GLLTEKTSSKVSADKHENSPG-IKVGVSGKKRRLESEQGKLAPRVDESSRAAKKPRCESA
HULK1    0.98 GCRKESSDSKVVTDLNIASSKKPKELLKEKKKRFENELGKSASGADESKRAAKRPRSEDA
HULK2    1.00 GSDEENASEIATVESDNN-RNEGNGVDSGSKVEQIDIGGKFLEGDYDLNKGLNFQINIMV
HULK3    0.99 GSDEDNASEIATVESNNNSRNEGNGVDSGSKVEYSDAVGEGCDGGHELNKGLDFHISTMV
         cons Gsdeenas   tv snnnsrneg gvdsg Kv  ede Gk a g de n   kfp sem

HUA2     0.99 DNKVKCEID------DG----SDSTGTVSDIKREIVLGLGARGGNFQYDKEAVAYTKRQR
HULK1    0.98 KDQKQCKSKRLVPVGEGKAEISDSTGVVSIFKREIVLGISALGGKNQFDKDMVAYTKRRK
HULK2    1.00 KRKKRKPTR--------KRGTSDVVDPQAKVEGEAVPEAGARNNVQTSQNSHEKFTERPC
HULK3    0.99 KRKKRKPTR--------KRETSDIIDPPAKVEAEGLGPN-ACDSCQRSQNSHERLNERPC
         cons krkkr ptr        kretSDst pv kv rEivlg gArgg qqs  sh ayt Rpc

HUA2     0.99 QTMEHATSPSFSGSRDKSGKGHLEQKDRSSPVRNVKAPAAQSLKKRRAVCIYDEDDD---
HULK1    0.98 QTVEHTSVSSFPGSLVKEGANHPEQKISSSSDSDIKVQAAQLPKRRRAVCIYDDDDDD--
HULK2    1.00 EENGDEHLPLVKRARVRMSRAFYGNHEANSSLQAEERSPKDTVVSATAQTSPSD-----I
HULK3    0.99 EENGDEHLPLVKRARVRMSRAFYADEKVNASSQVEERSSKDTLLSAALQTSPSVNHENGI
         cons   n  ehlp  k  rv m ra yeqk   ss q e rsa  tlks ra     dddd

HUA2     0.99 -EDPKTPLHGKPAIVPQAASVL-TDGPKRANVCHSTSTKAKISAGSTESTEVRKFPLRKH
HULK1    0.98 -EDPKTPVHGGLSNIPIAS----TDAPKSANASHNTSIKAKLLAGSTDSVKTGKVPLYKH
HULK2    1.00 ISSHDTFAVEESKFFEVSAKLSGDMVNVAPSPVEKSHDGMSPSEACVQTVREREYAMGW-
HULK3    0.99 GSGHDTSAAEEFNSFELSAKLSGVMVDVVPSHMEKPSDRMSPSVACVQTVGDRQTAVNFH
         cons   d  Tpah e   f   akls t vp       ktsdk  psa   q v  rk  l kh

HUA2     0.99 CEDASRVLPS------------------------------------------------
HULK1    0.98 NKDASLALPDSVEGYNSRMGKPFKALLQKNIKPILRSPKNSYQ---LVSFKKQVTGQN--
HULK2    1.00 -NELSKTPDDKSAGPQYNQVSSLPAGEAQTASVPEAVCPEVLK--LLTSESDLPAVQY-C
HULK3    0.99 ENEFTMTLDDEVTRAQSNQLSSLVETEARVPEVVQGCSEESQTGNCLISETDPIDIQCSH
         cons  n as tl d v g qsnq ssl a ea    v      es     l se d    q
```

```
HUA2     0.99 ------------NAENSTNSLPVVKPINELPPKDVKQILQSPKMSPQLVLTNKHVAGQHK
HULK1    0.98 KTAKVAGAGMPDSVEGPSNSSYMGKPVIKLPPQNVKQTLRSPKKSPQLFSTKELVAVQNK
HULK2    1.00 QVAKIEPSMDPNTVDSSANNASEICS----LS--IPSQLSGQDRSNDQD-ACVSLENSRE
HULK3    0.99 QSEKHETPLNPDIVDSSANKSPGLCSSLDMTTTVVPAQSPHQHKIQEYDSSDHSLVIVGD
         cons q ak e    pd v ssaNssp       lpp  v qql s kkspqldst  s a q k

HUA2     0.99 VVKSSVKVSGVVMAKKPQSDSCKEAVAGSDKISSSQSQPANQRHKSASVGERLTVVSKAA
HULK1    0.98 IAK----VSGAGIPKKYHGDSSKDVVAGSDRVSSSHSQTANQRSK-PAFGEKPTSTPKVA
HULK2    1.00 YLNEEGSKIDACVAQVVQSEAIEHSP-SSCLVVNKQETENMPKTVNMLLKEGHGSLGEEC
HULK3    0.99 SLNGKCEKIDYCMTQVVQSQALEPPPPLFCSVVNYQEVENLQETENTLWKENQGSPGKEL
         cons  l        acma  vqsd      ags  v  sq qe nqrtkn l  E   s gkea

HUA2     0.99 SRLNDSGSRDMSEDLSAAMLDLNREKGSATFTSAKTPDSAAS---MKDLIAAAQAKRKLA
HULK1    0.98 TRLDVEVSRDTFVNLSADVIDVNQENGNAPLFSFGMSDSSSSC--MKDLIAAAQAKRKQA
HULK2    1.00 AIVEPAQCTPNLPISATESDVIVGENVPLNEIGCTKCEDAVEDSRQLKMIGETNDQKQQV
HULK3    0.99 DSDKQAHMIQNPVLSATESEMIVDDAEPQYETVYSHCADAVEN-RELEKSCEVDEQKEQM
         cons  rl  a srdn v    es di  engpa ets   cd av    m dlia aqa  kqa

HUA2     0.99 HTQNSIFGNLNPSFLSISDTQGR--SHSPFMVQNASASAAISMPLVVQGHHQQGSSPSNH
HULK1    0.98 HSQFSPFVNLDHNSLNIDSMQT---SKSPFMVQNVSSPAADAT-LIVAQEHQEVLTPSNH
HULK2    1.00 QTNNSVLVSENLSREKMSFSPAITADTPARGTPHSSSVYYHISTSESANDMQNNSSGSPN
HULK3    0.99 QATNSISVSENFSREKLNSS-------PARGTPNCNSVCR-ISTAESENAMQNNSYYSTN
         cons  tqnSifv  n sr kisssq    s       n ssvaa istle an  QnnsspSn

HUA2     0.99 GHQSLSRNQIETDDNEERRLSSGHKSVGGSLSCSTEAAISRDAFEGMLETLSRTRESIGR
HULK1    0.98 GRQSSSSNQAGTEENEERRFSSGHRSVGGSLSGATEAAISRDTFEGMIETLSRTKESIRR
HULK2    1.00 IPTGEKKNDCDAIVKEEEKIETGVCQGQKVVS--CDVQSTRESYEDALCSLVRTKESIGR
HULK3    0.99 VQYGENKSLNVDTVKEESKVETGTTQVKKVVS--SDVQCTVESFETALDSLVRTKETIGR
         cons g q esknq  t v EEr    Gh  vg   S  t   i r sfEg le L RTkEsIgR

HUA2     0.99 ATRLAIDCAKYGLASEVVELLIRKLESESHFHRKVDLFFLVDSITQHSHSQKGIAGASYV
HULK1    0.98 ATRVAIDCAKYGIANEVVELLIRKLEIEPHFPRKVDLFFLLDSIIQSSHSQKGRARSLYI
HULK2    1.00 ATCLAMDLMKFGVSAKAMEILAHTLESESNLKRRVDLFFLVDSIAQCSKGLKGDTGCVYL
HULK3    0.99 ATRLAMDLAKFGVSAKAMEILAHTLESESNLQRRVDLFFLVDSIAQCSKGLNGDAGGVYL
         cons ATrlA D aK Gv a   E L   LEsEs   R VDLFFLvDSIaQcS   kGdag vYl

HUA2     0.99 PTVQAALPRLLGAAAPPGTGASDNRRKCLKVLKLWLERKVFPESLLRRYIDDIRASGDDA
HULK1    0.98 PTVQAALPRLLGAAAPPGTGARENRHQCRKVLRLWLKRKIFPDFLLRRYIGDLGASGDDK
HULK2    1.00 SAIQVILPRLLAAAVPAGATTQENRKQCLKVLKLWLERRILPESIVRHHIRELDSHS-IV
HULK3    0.99 SSIQAMLPRLLTAAVPAGATTQENRKQCLKVLRLWLERRILPESIVRHHIRELDSLS-NV
         cons  t QaaLPRLLgAA P G   qeNRkqClKVL LWLeR i Pes  R  Ir ld s  dv

HUA2     0.99 TGGFSLRRPSRSERAVDDPIREMEGMLVDEYGSNATFQLPGFFSSHNFEDDEEDDDLPTS
HULK1    0.98 TVGFSLRRPSRSERAVDDPLRDMEGMLVDEYGSNANFQLPGYLASLTFGDDEEEDLPSTS
HULK2    1.00 PACLYSRRSARTERSLDDPVRDMEDMLVDEYGSNSTLQLPGFCMPALLKDEEGGSDSEGG
HULK3    0.99 PACLYSRRSARTERALDDPVRDMEGILVDEYGSNSTLQLHGFCIPPILRDEDEGSDS---
         cons  a    RR  R ERa DDPvRdMEgmLVDEYGSN t QLpGfc      D eeg ds ts

HUA2     0.99 QKEKS----------TSAGERVSALDDLEIHDTSSDKCHRVLEDVDHELEMEDVSGQRKD
HULK1    0.98 QEVKNT---HMEVKITHMEEPVLALGKLEAHDSSSDKPH-CVVDVNGGLEMEDASCQLKD
HULK2    1.00 CDSEGGSDSDGGDFESVTPEHESRILEENVSSSTAERHTLILEDVDGELEMEDVAPPW--
HULK3    0.99 ---------DGGDFESVTPEHESRSLEEHVTPSITERHTRILEDVDGELEMEDVAPPWEG
         cons q  k     dggdfe vtpEh s lle evhdsss  h rileDVdgeLEMEDv p wkd

HUA2     0.99 VAPSSFCENKTKEQSLDVMEPVAEKSTEFNPLPEDSPPLPQESPPPLPPLPPSPPPPSPP
HULK1    0.98 ---D-VCGIEAKED-----SPATTCATELPSFPAGSPPLPHESPPSPPPQPPSSPPP-PS
HULK2    1.00 --GTENCTHTDQAD------NTKVSNCQLGQQHRPVFGTSHQHMSLSSPPLPSSSPP---
HULK3    0.99 --GSSASAITDQAD------NRESANCLL------VPGTSHQNVTSSSP-----------
         cons   gss c itd  d           n el   p   p   h sppss P ppssppp
```

```
HUA2     0.99 LPPSSLPPPPPAALFPPLPPPPSQPPPPPLSPPPSPPPPPPPPSQSLTTQLSIASHHQIP
HULK1    0.98 SPPQLAPAPPPSDHCLP-------PPTAPLAPAQSIALPP-----SSITRPSMPSHPSLP
HULK2    1.00 ------------------------PPPAPPSQQGECAMPD-------------------S
HULK3    0.99 --------------------------PARPSQNAQLAMSN-------------------S
         cons                        pppap s   s ampp

HUA2     0.99 FQPGFPPPTYPLSHQTYPGSMQQDRSSIFTGDQIVQGPGNSSRGGLVEGAGKPEYFVQQS
HULK1    0.98 LQPGFAPPAYPLLQHEYQISMQRDHSSIATSNQIAPVPVNAAHGRHADGGVKSEYLMPQS
HULK2    1.00 YLNGFENGGYRNVHGDQQAGPLRMNPPLSGSTMHYQGPESSYISGVQLTNSIP---QADG
HULK3    0.99 YSNGFDYRRNPSMQGDYHAGPPRMNPP-----MHYGSPEPSYSSRVSLSKSMP---RGEG
         cons yq GF pp ypl  gdyqa  qr n  i ts   yqgPensy   v lg skp     q

HUA2     0.99 SSFSPAGVCSSR----EPSSFTSSRQLEFGNSDVLFNPEASSQNHRFQPS-TPLSQRPMV
HULK1    0.98 SSFAPVGMCSYG----EPLPFISSKQLEYGNSDVLFKQEASSQNQQLRPINTSFLQRPMI
HULK2    1.00 SNFQHRPYPSHPHPHPPPPPPPPQHQFSFREPGHVLKSHRDAPSYSHRSH----------
HULK3    0.99 SNFQHRPYPSS-----HPPPPPPSHHYSYMEPDHHIKSRREGLSYPHRSH----------
         cons S Fq r y Ss     ePpp p shql  g  d lfkse ssq y hr h

HUA2     0.99 R-LPSAPSSHFSYPS-HIQSQSQHS-YTHPYPFPPQRDDARRYRNEEPWRIPSSGHSAEN
HULK1    0.98 RNLAPAPSSHFPLPCRIVQSEPQRSSFPHPYHFPSQPVDGRQHMN-EEWRMPPNGCSADP
HULK2    1.00 -YVPNCDERNFHDNHERMRHAPFEN--RDNWRYPPSSSYGSRYQDEHKAPYPSSSYNG--
HULK3    0.99 -YTLEFDERNYQDSYERMRPEPCEN--RDNWRYHPPSSHGPRYHDRHKGPHQSSSYSGHH
         cons  ylp a    f dp erm sepqe   r   r ppqssdgrry  e kw  pss ys

HUA2     0.99 QNGAWIHGRNSHPGLPRVTDSFFRPPPERPPSGTMNYQPSAASNLQAVPAIPGHTAPQML
HULK1    0.98 QYGAWIGVRNPFPGSRTVTDGVFQPPPERPPSGTVRYQ-LAANNLQGGSTISGNIASQML
HULK2    1.00 -----VR-------------------WDNPPRQYNNRPSFHPKPHSEGPAPVGMRDPGMW
HULK3    0.99 RDSGRLQNNR----------------WSDSPRAYNNRHSYHYKQHSEGPVPVGMRDPGTW
         cons q gawi  rn                erpP g nn qs  akn  egpa vGmr p m

HUA2     0.99 PSRPDIPTVNCWRPA
HULK1    0.98 LSRPDVPSAAQYRPS
HULK2    1.00 HQRSD----------
HULK3    0.99 HQR------------
         cons h Rpd
```

Figure 3.19. Subcellular localization of HUA2 in Agrobacterium-infiltrated tobacco leaves. Fluorescence and bright-field images of tobacco epidermal leaf cells infiltrated with a mixture of Agrobacterium suspensions harboring constructs encoding HUA2, the silencing suppressor p19 and mCherry-RFP (which served to identify the nuclei of tobacco epidermal cells). Tobacco leaves infiltrated with pBIN-attr-YFP as vector control **(A)** mCherry **(B)** or with *HUA2* **(C)** were analysed by confocal laser scanning microscopy 2 days after infiltration. Bright field images are indicated as Brightfield, GFP signal (green) is indicated as GFP and the RFP signal (red) which served as nuclear marker is indicated as RFP. The last panel of **(C)** shows a merged image (indicated as overlay) Scale bar: 25µm.

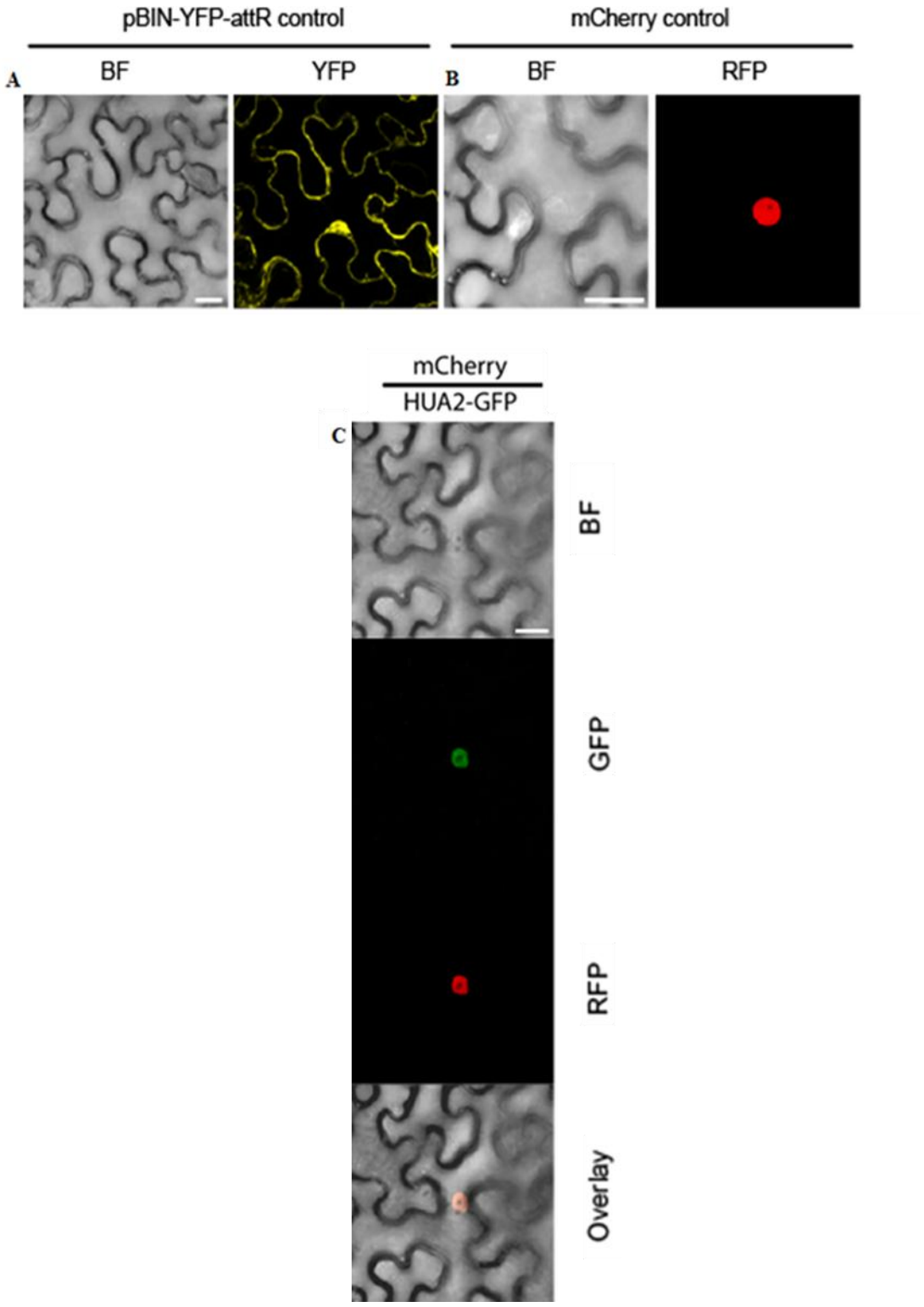
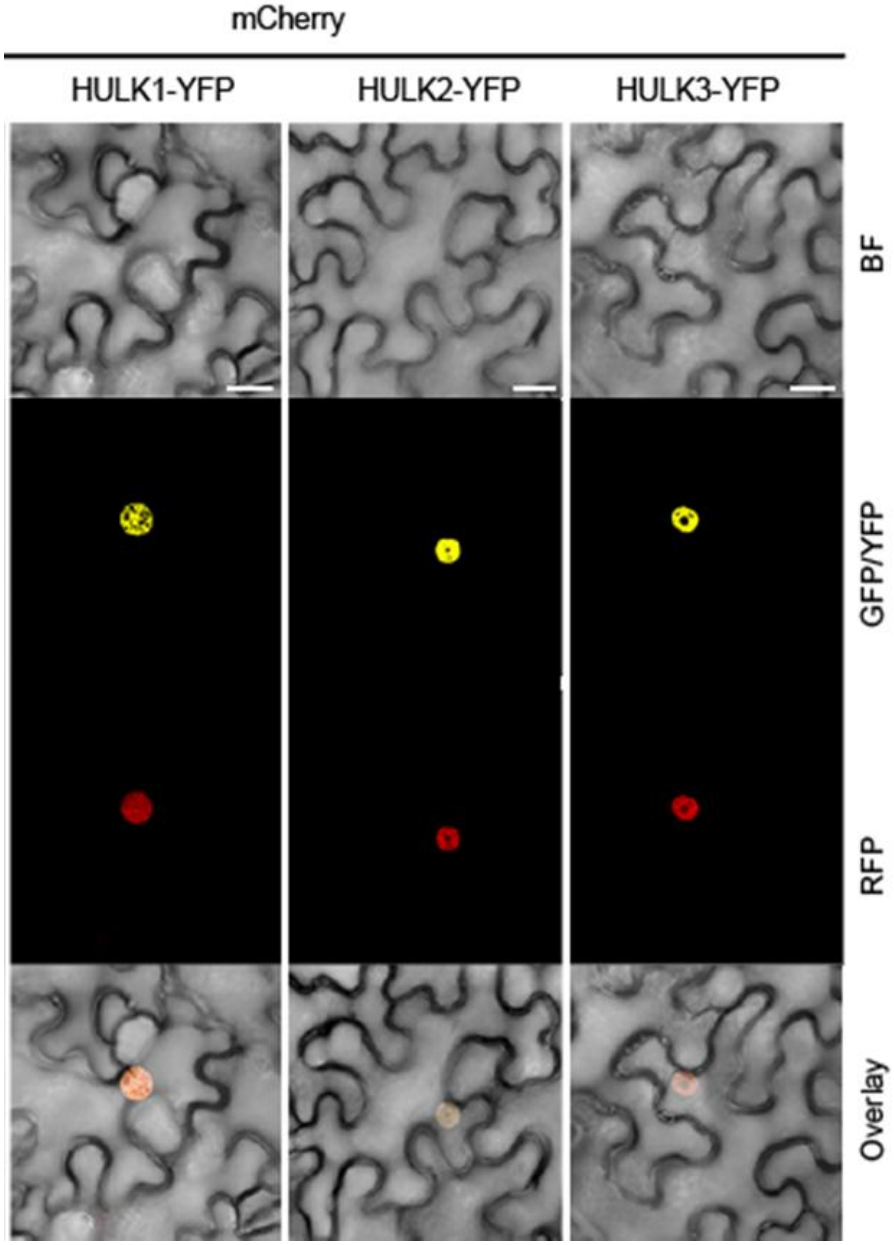

pBIN-YFP-attR control
mCherry control
A
BF
YFP
B
BF
RFP
mCherry
HUA2-GFP
C
BF
GFP
RFP
Overlay

Figure 3.20. Subcellular localization of HULK1, HULK2 and HULK3 in Agrobacterium-infiltrated tobacco leaves. Fluorescence and bright-field images of tobacco epidermal leaf cells infiltrated with a mixture of Agrobacterium suspensions harboring constructs encoding the gene of interest, the silencing suppressor p19 and mCherry-RFP (which served to identify the nuclei of tobacco epidermal cells). Tobacco leaves infiltrated with genes of interest were analysed by confocal laser scanning microscopy 2 days after infiltration. Bright field images are indicated as Brightfield, YFP signal (yellow) is indicated as YFP and the RFP signal (red) is indicated as RFP. The last row shows the merged images (indicated as overlay) Scale bar: 25µm.


mCherry
HULK1-YFP
HULK2-YFP
HULK3-YFP
BF
GFP/YFP
RFP
Overlay

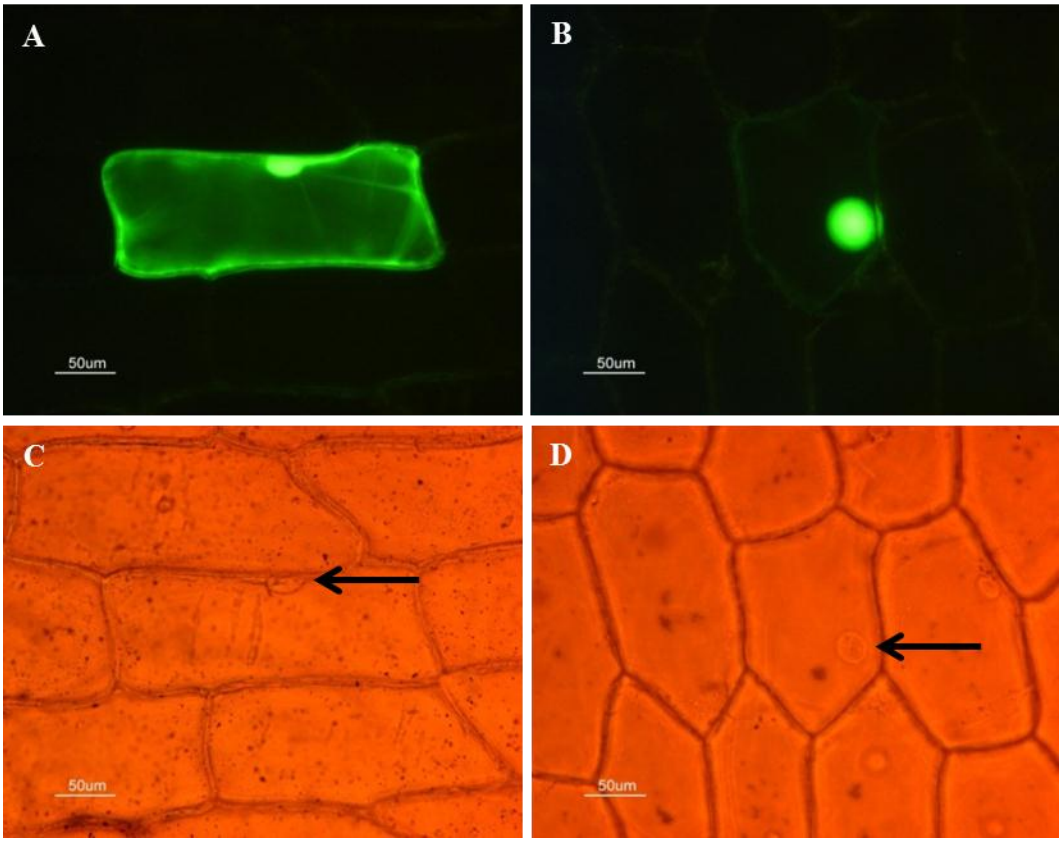
of tobacco leaf epidermal cells (Figure 3.19). Similarly HULK1, HULK2 and HULK3, were all observed to localize within the nucleus of tobacco cells (Figure 3.20). These data confirmed that the HUA2 protein family members do indeed localize within the nucleus of the plant cell.

3.7 The amino terminal region is required for restricting HUA2 within the nucleus

In silico analysis using the program NucPred predicted that the NLS within HUA2 is present in the amino terminal region. To confirm this I performed *in vivo* targeting experiments in tobacco (*Nicotiana benthamiana*) using agroinfiltration, as well as in onion (*Alliumcepa*) epidermal cells using particle bombardment. To determine the intracellular localization of amino terminus portion of HUA2 (NT-HUA2), a construct was made in which NT-*HUA2* was fused to *GFP* and was placed under the control of the CaMV 35S promoter. The construct was introduced into onion epidermal cells using particle bombardment. In cells where the empty GFP vector was introduced alone (Figure 3.21 A) fluorescence was observed throughout the nucleus and cytoplasm. However, in cells expressing 35S:*NT-HUA2-GFP*,fluorescence was observed only in the nucleus (Figure 3.21 B). The position of the nuclei in A and B was determined under bright field illumination and is shown in Figure 3.21, panels C and D (arrows).

Next, I determined the localization of CT-HUA2 within the cell. To this end, a 35S:CT-HUA2-GFP construct was generated and delivered into tobacco leaf cells by *Agrobacterium tumefaciens* infiltration (de Felippes and Weigel, 2010). 35S:CT-HUA2-GFP was co-expressed with mCherry-NLS (a protein known to localize in the nucleus)

Figure 3.21. Subcellular localization of NT-HUA2-GFP after transient transformation of onion epidermal cells. Cells were visualized using florescence microscopy: **(A)** Onion cells transformed with empty GFP vector and **(B)** Onion cells transformed with construct encoding NT-HUA2-GFP. The position of the nuclei in **(A)** and **(B)** was determined under bright field illumination and is shown in **(C)** and **(D)** (black arrows).


A
50um
B
50um
C
50um
D
50um

and p19 (a protein known to supress gene silencing) (Karimi et al., 2002; Shaner et al., 2004; Voinnet et al., 2000; Harter, unpublished). Strong GFP signals were visualized throughout the tobacco cell and were not restricted to the nucleus (Figure 3.22).These sub-cellular localization results indicate that the amino terminus end of HUA2 contains an NLS and that it is required, as predicted, to restrict HUA2 to the nucleus of the plant cell.

3.8 HUA2 interacts with the *Arabidopsis* CTD independently of the DSI motif within its RPR domain in yeast cells

The C-Terminal Domain (CTD) of RNA pol II plays an important role in transcription through the recruitment of various factors required for transcript initiation, elongation and/or maturation. The elongation of a transcript is a tightly regulated event in which trimethylation of H3K4 at the transcription-start site of an active gene is required for RNA Pol II to proceed with elongation. In *Arabidopsis*, this methylation state is affected by the Paf1 complex consisting of *ELF7*, *ELF8* and *VIP4* which is recruited to the transcription-start site via interaction with the CTD (He et al., 2004). Our laboratory has previously demonstrated that along with components of the Paf1 complex, *HUA2* is also required for trimethylation of histone 3 at the *FLC* locus (Sajja and Grbic, unpublished). The RPR domain within HUA2 has DSI motif similar to other CTD binding proteins such as the *S. cerevisiae* Nrd1p and Pcf11 (Steinmetz and Brow, 1998; Barillà et al., 2001; Sadowski et al., 2003). The RPR domain has a small motif containing three highly conserved amino acids DSI (Figure 3.23). It has been shown that an intact DSI motif is required for the RPR domain of Pcf11 to interact with the CTD of RNA Pol II (Sadowski et al., 2003).

Figure 3.22. Subcellular localization of CT-HUA2 in Agrobacterium-infiltrated tobacco leaves. Fluorescence and bright field images of epidermal leaf cells infiltrated with a mixture of Agrobacterium suspensions harboring constructs encoding CT-HUA2, the silencing suppressor p19, and mCherry-RFP (which served to identify the nuclei of tobacco epidermal cells). Tobacco leaves infiltrated with the genes of interest were analysed by confocal laser scanning microscopy two days after infiltration. Bright field images are indicated as BF, GFP signal (green) is indicated as GFP and the RFP signal (red) as RFP. Overlay shows the merged image.

CT-HUA2-GFP+mCherry-RFP

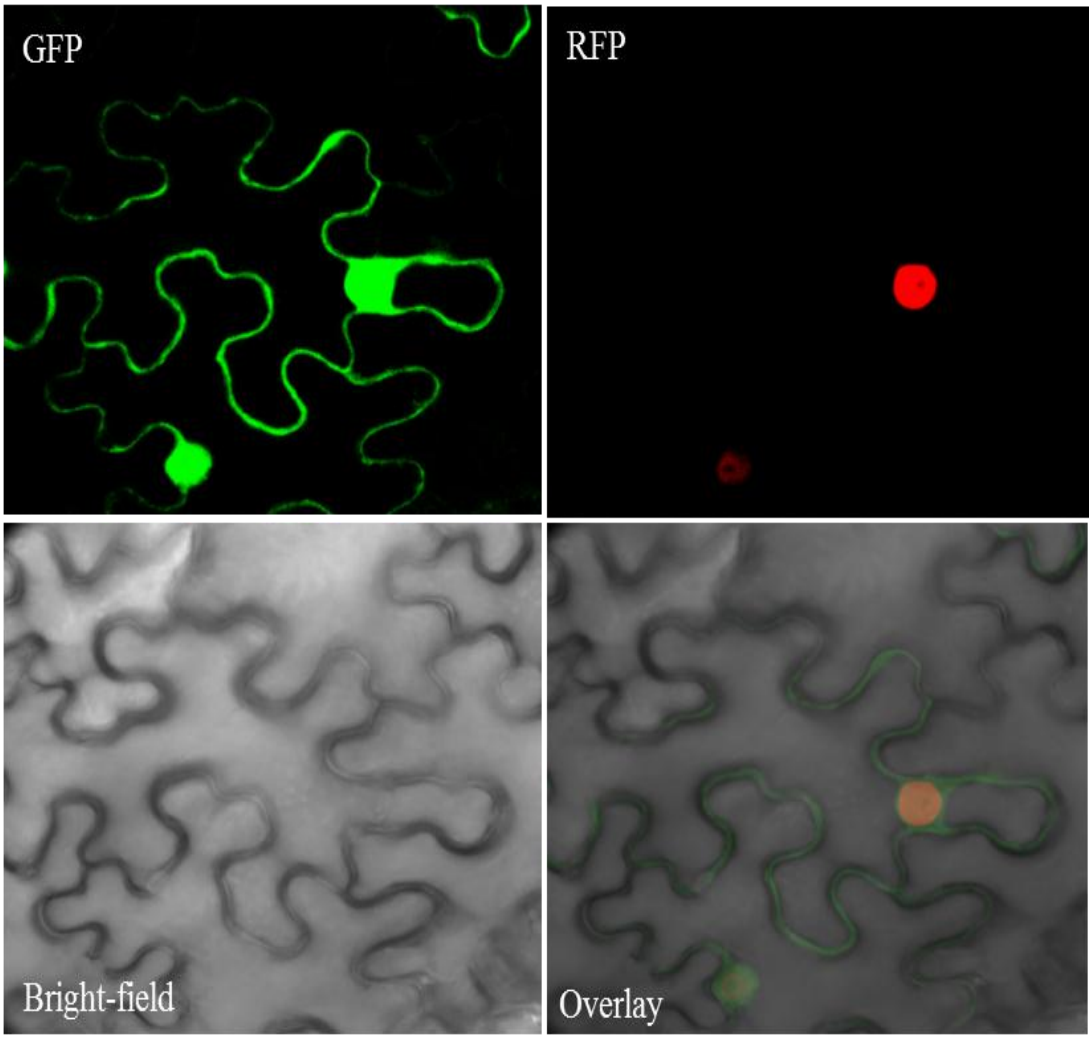
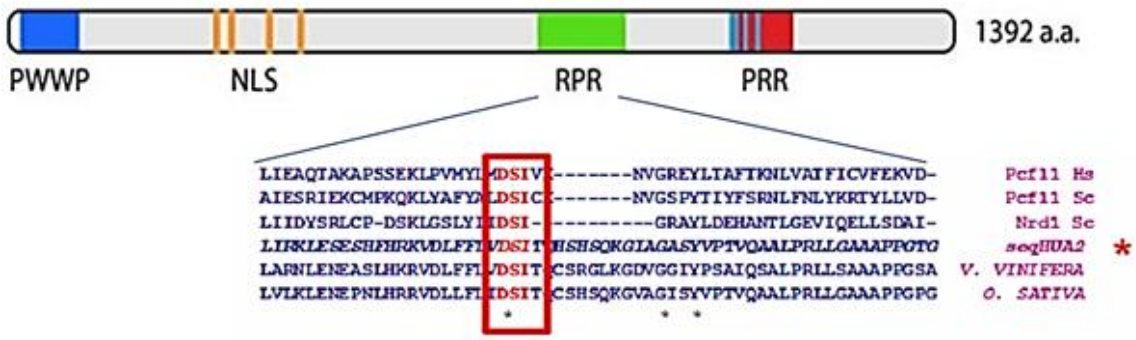


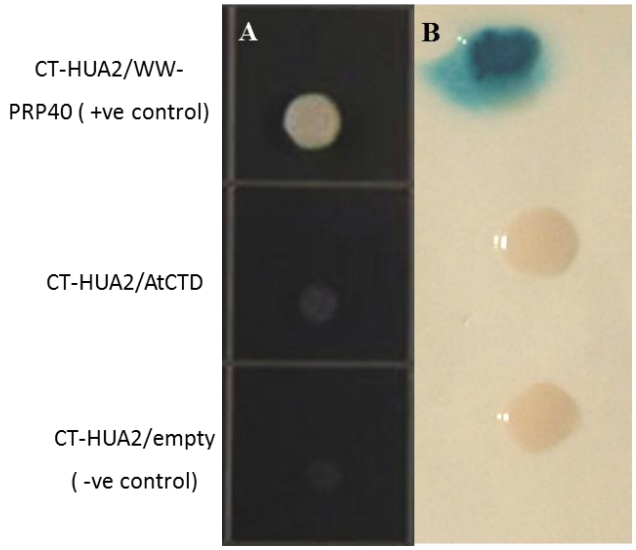
Figure 3.23. ClustalW alignment of the RPR domain from *Saccharomyces*, *Arabidopsis*, *Oryza* and *Vitis*. The conserved motif DSI is shown in red in all the four species. Asterisk represents the RPR domain sequence from *HUA2*.


1392 a.a.
PWWP
NLS
RPR
PRR
Pcf11 Hs
Pcf11 Sc
Nrd1 Sc
seqHUA2
V. VINIFERA
O. SATIVA

This led us to test if HUA2 interacts with the *Arabidopsis* CTD (AtCTD) and furthermore if this interaction is dependent on the DSI motif. To this end, I carried out yeast two-hybrid analysis to determine if AtCTD interacts with (1) C-terminal region of HUA2 lacking the RPR domain ,(2) full-length HUA2 and (3) the HUA2ΔRPR mutant in which the highly conserved DSI motif of the RPR domain is replaced by three consecutive alanines. I co-transformed MVa203 yeast cells with a plasmid encoding the GAL4 DNA-binding domain fused to CT-HUA2, HUA2or HUA2ΔRPR, together with a plasmid encoding the GAL4-AD-AtCTD fusion and then assayed for activation of *HIS3* and *Lac*Z transcription. Figure 3.24(upper panel A and B) shows that CT-HUA2 which lacks the RPR domain does not show interaction with AtCTD as indicated by no growth on –HIS medium, on the other hand, both full-length HUA2 and the RPR mutant HUA2ΔRPRinteract with AtCTD as indicated by the growth on –HIS media under stringent conditions (80 mM and50 mM 3AT respectively) (Figure 3.24 A lower panel). X-gal filter assay (Figure 3.24B lower panel, Table 3.8). This suggests that interaction between HUA2 and AtCTD requires protein domains upstream of CT-HUA2 sequences. In addition I carried out an inverse yeast two-hybrid assay where I switched the bait and prey proteins. For this I created a construct encoding AtCTD fused to the GAL4 DB and checked its interaction with HUA2 or HUA2ΔRPR (fused to the GAL4 AD). I first determined the auto-activation of GAL4 DB-AtCTD and determined that a concentration of 110mM 3AT is sufficient to supress GAL4 DB-AtCTD auto-activation (Figure 3.25). For this assay I co-transformed MVa203 yeast cells with plasmids encoding GAL4 DB-AtCTD and GAL4 AD-HUA2 or GAL4 AD-HUA2ΔRPRand assayed for activation of *HIS3* and *Lac*Z transcription.

Figure 3.24. Yeast two-hybrid analysis to determine the interaction between CT-HUA2, HUA2 and HUA2ΔRPR with AtCTD. **(A)** MVa203 cells were co-transformed with the indicated constructs and grown on selection plates lacking leucine, tryptophan and histidine (-LTH) with 50 mM 3-AT 2. **(B)** X-gal filter assay for colonies show in **(A)**.

GAL4DB/AD fusions



GAL4DB/AD fusions

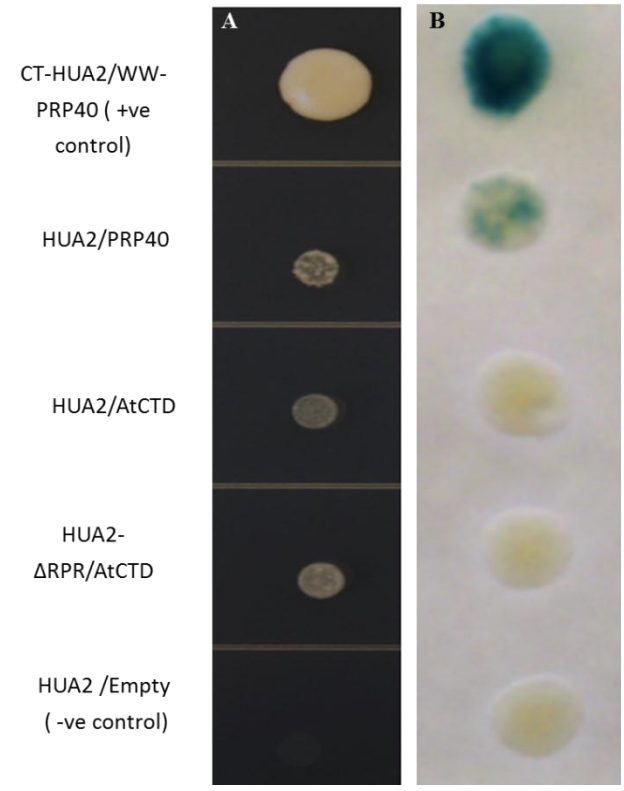


Table 3.8. Summary and interpretation of yeast two-hybrid assays conducted with the indicated bait and prey proteins.

Fusion gene with				
GAL DB	*GAL*AD	*HIS*3 Expression	β-galactosidase Expression	Interpretation
CT-HUA2	*PRP40* (P)	+	Blue	Strong interaction
CT-HUA2	*AtCTD*	-	White	No interaction
HUA2	*PRP40*	+	Blue	Strong interaction
HUA2	*AtCTD*	+	White	Weak interaction
HUA2ΔRPR	*AtCTD*	+	White	Weak interaction
HUA2	Empty(N)	-	White	No interaction

Figure 3.25.Determination of the optimal concentration of 3AT for assaying GAL4 DB-PRP40(WW) and GAL4 DB-AtCTD fusions. The concentration of 3AT at which the basal expression of reporter histidine gene is supressed was determined for cells expressing GAL4 DB-PRP40(WW) and GAL4 DB-AtCTD fusions. MVa203 cells were transformed with a plasmid encoding GAL4 DB:PRP40(WW)and GAL4 DB:AtCTD and then plated on media plates lacking leucine and histidine and carrying the indicated concentration of 3AT. No growth was observed at 20 mM 3AT for GAL4 DB-PRP40(WW).

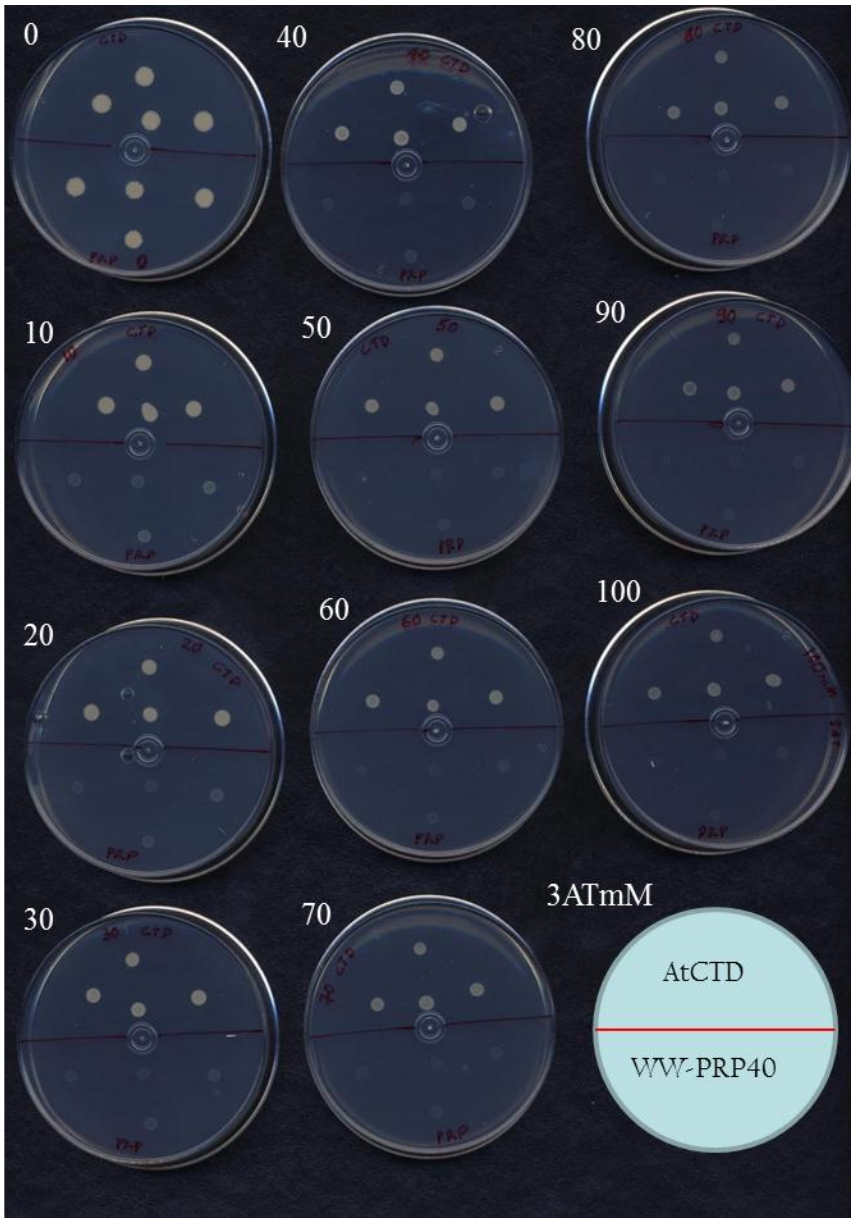

0
40
80
10
50
90
20
60
100
30
70
3ATmM
AtCTD
WW-PRP40

Figure 3.26. Determining the optimal concentration of 3AT for assaying GAL4 DB:AtCTD fusion (continued). The concentration of 3AT at which the basal expression of reporter histidine gene is supressed was determined for cells expressing GAL4 DB-AtCTD. MVa203 cells were transformed with a plasmid carrying *GAL4 DB-AtCTD* and then plated on plates lacking leucine and histidine carrying the indicated concentration of 3AT. No growth was observed at 110 mM 3AT for GAL4 DB:AtCTD.

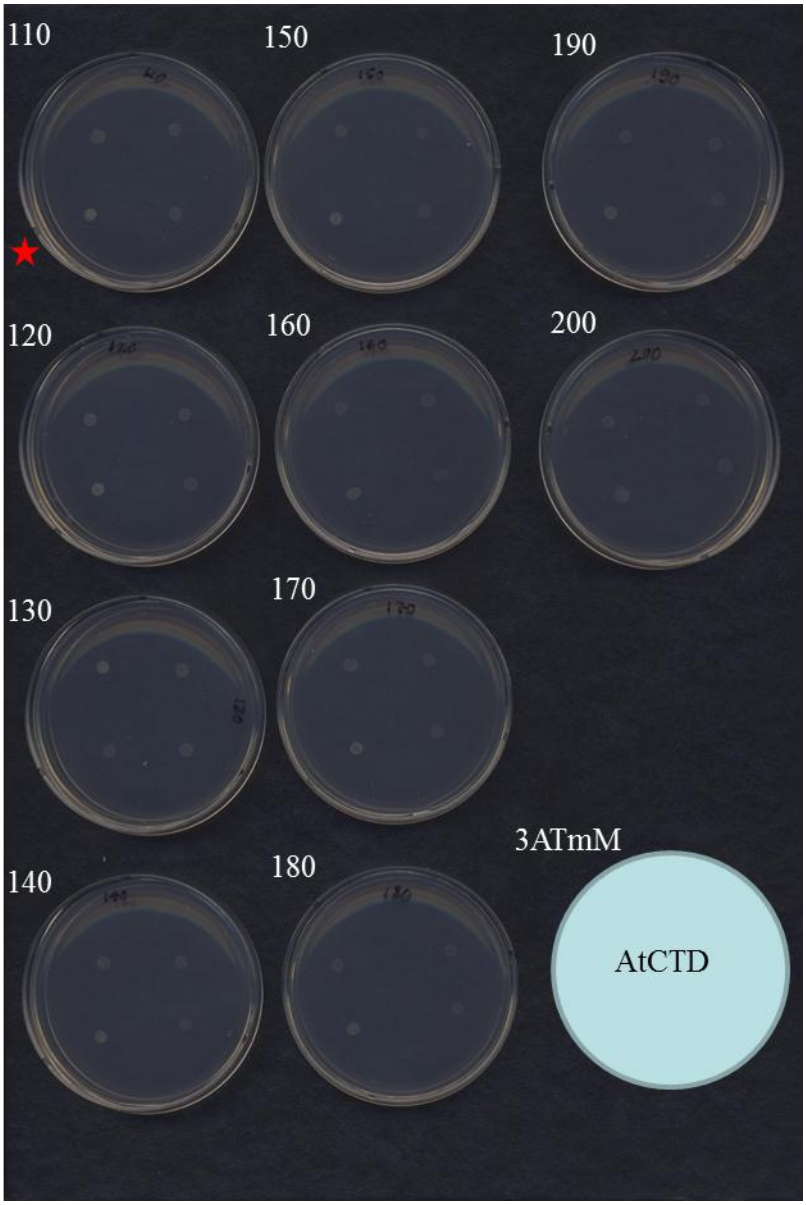

110
150
190
120
160
200
130
170
140
180
3ATmM
AtCTD

Figure 3.27A shows that AtCTD interacts with both HUA2 and HUA2ΔRPR and these results were confirmed by an X-gal filter assay (Figure 3.27 B). These data suggest that the disruption of the DSI motif within the RPR domain does not affect interaction between HUA2 and AtCTD in yeast cells.

3.9 The HULK protein also interacts with AtCTD in yeast cells

As mentioned earlier HUA2 and the HULKs share the same conserved domain structures and significant amino acid sequence similarity. All the HULKs also contain a putative RPR domain which has been suggested to be involved in interacting with CTD of RNA Poll II. I thus further carried out yeast two-hybrid assays to see if the HULKs also interact with AtCTD. I co-transformed MVa203 cells with a plasmid expressing the GAL4 DB- AtCTD fusion together with a plasmid expressing the GAL4- AD fused to full length *HULK1*, *HULK2* or *HULK3*. Figure 3.28 shows that AtCTD interacts with all three HULKs in yeast cells (Table 3.9).

3.10 The binding of HUA2 to the CTD requires phosphorylation of Ser-2 in the heptad repeats

The CTD of RNA Pol II is composed of tandem repeats of the sequence YSPTSPS. These repeats control the recruitment of factors that promote RNA synthesis and processing. This recruitment and binding of transcription factors to the CTD depends on differential phosphorylation of Ser-2and Ser-5 residues within the heptad repeats(Buratowski, 2005; Phatnani and Greenleaf, 2006; Saunders et al., 2006). After observing the interaction between HUA2 and AtCTD, I was curious to determine if the interaction between HUA2 and the CTD depends on CTD phosphorylation state.

Figure 3.27. Yeast two-hybrid analysis of the interaction between AtCTD and HUA2 or HUA2ΔRPR. (A) MVa203 cells were co-transformed with constructs encoding the indicated constructs and grown on selection plates lacking leucine, tryptophan and histidine (-LTH) with 120 mM 3AT. **(B)** X-gal filter assay for colonies shown in **(A)**.

GAL4DB/AD fusions

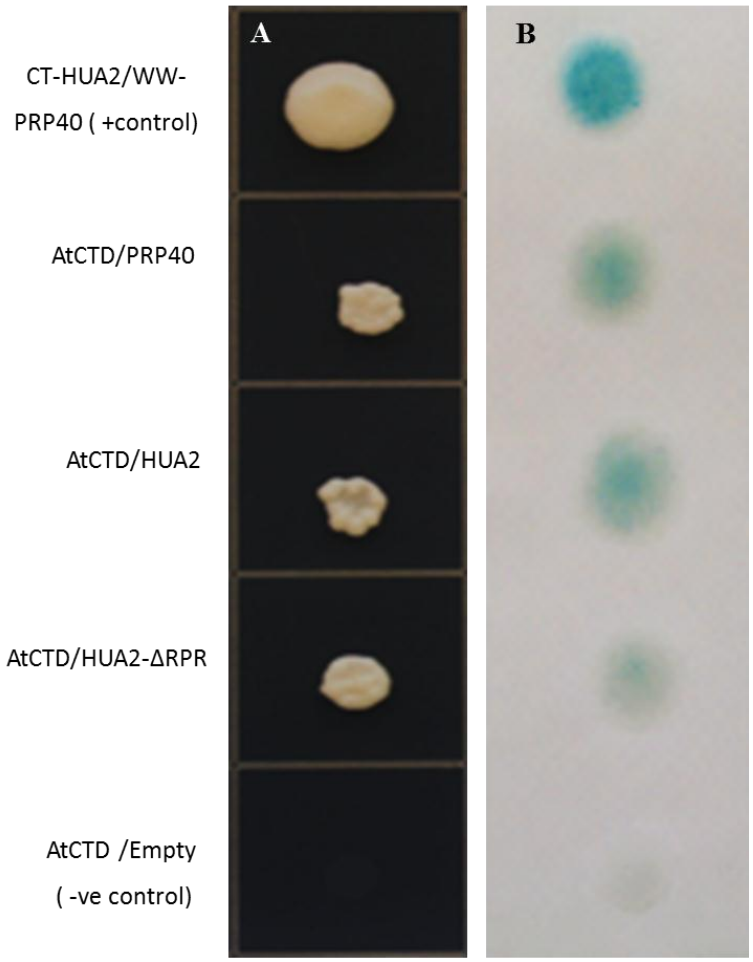


Figure 3.28. Yeast two-hybrid analysis of the interaction between AtCTD and HULK1, HULK2 or HULK3. **(A)** MVa203 cells were co-transformed with constructs encoding the indicated fusion proteins and were grown on selection plates lacking leucine, tryptophan and histidine (-LTH) with 120 mM 3-AT. **(B)** X-gal filter assay for the colonies shown in **(A)**.

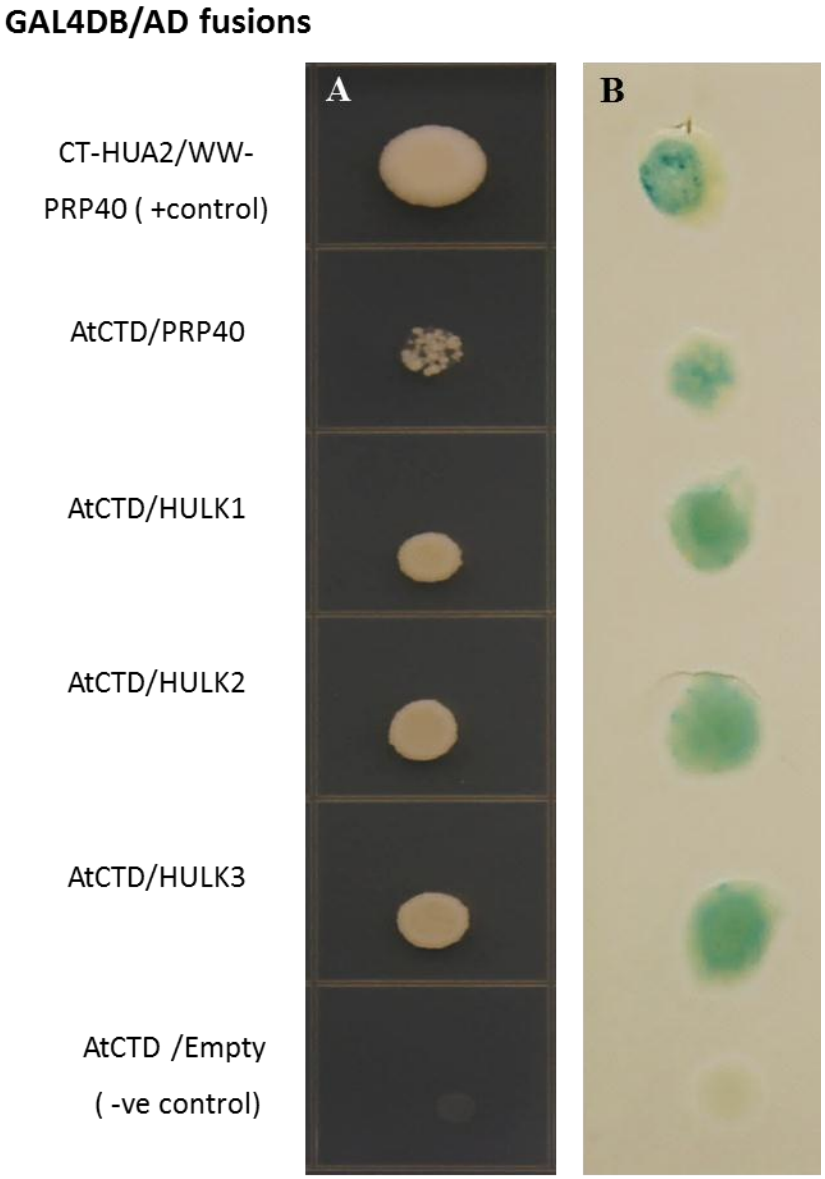

GAL4DB/AD fusions
A
B
CT-HUA2/WW-PRP40 (+control)
AtCTD/PRP40
AtCTD/HULK1
AtCTD/HULK2
AtCTD/HULK3
AtCTD /Empty (-ve control)

Table 3.9. Summary and interpretation of yeast two-hybrid assays conducted with the indicated bait and prey proteins.

Fusion gene with				
*GAL*DB	*GAL*AD	*HIS*3 Expression	β-galactosidase Expression	Interpretation
CT-HUA2	*PRP40* (P)	+	Blue	Strong interaction
AtCTD	*HUA2*	+	Blue	Strong interaction
AtCTD	*HUA2ΔRPR*	+	Blue	Strong interaction
AtCTD	*HULK1*	+	Blue	Strong interaction
AtCTD	*HULK2*	+	Blue	Strong interaction
AtCTD	*HULK3*	+	Blue	Strong interaction
AtCTD	Empty (N)	-	White	No interaction

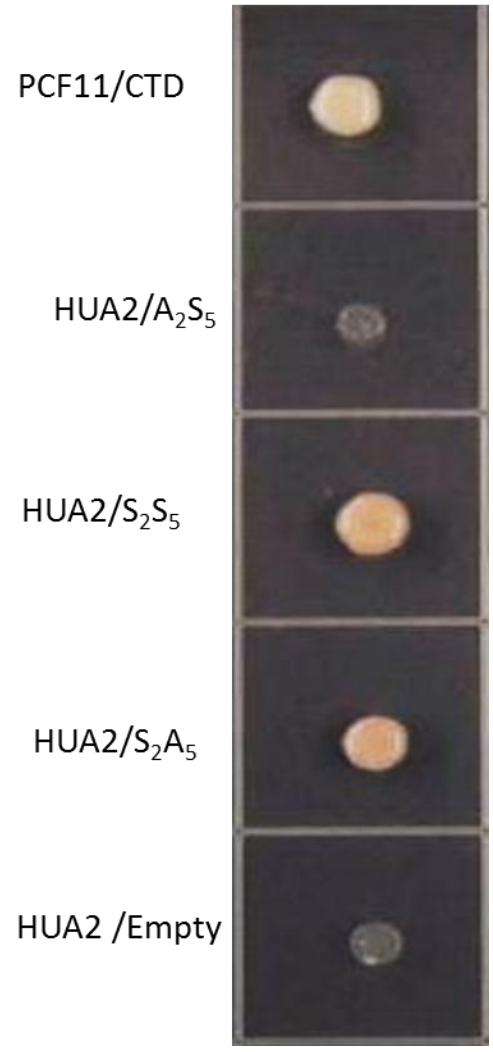
To determine this, I made use of a modified two-hybrid system described by Ursic et al., (2008) that detects interactions between a protein of interest and different phosphorylated forms of the yeast CTD RNA Pol II (CTD) . The constructs for performing the assay were obtained from Dr. Culbertson from The University of Wisconsin. Three different constructs encoding different CTD fusions were used: one encoding serine residues at positions 2 and 5 (pS_2S_5), one encoding alanine substitutions at position 2, (pA_2S_5), and one encoding an alanine substitution at position 5 (pS_2A_5). The substitution of alanine for serine prevents phosphorylation at that substituted position. To perform the two-hybrid assay, constructs encoding the GAL4DB-HUA2 fusion was co-expressed with the cells expressing GAL4AD-CTD fusion variants. Figure 3.29 shows that HUA2 interacts with S_2S_5 and S_2A_5 fusions, but not with the A_2S_5. This suggests that phosphorylation of Ser-2 is a requirement for interaction between HUA2 and the yeast cells.

3.11 Significance of the RPR domain of HUA2 using genetic complementation

After learning that the RPR domain containing portion of HUA2 is required for interaction with AtCTD in yeast cells, I was curious to determine the functional significance of the RPR domain within HUA2 *in planta*. Therefore, I performed a complementation test using *FRI-hua2-3* and *hua2-7/hulk1* double mutants. *FRI hua2-3* is an early flowering genotype relative to *FRI-Col-0*. The loss of function of *HUA2* in *FRI* Col-0 (*FRI hua2-3*) makes them early flowering. The *hua2-7/hulk1* double mutant is slightly early flowering but shows prominent embryo defects within the siliques in later stages of development relative to *hulk1* and *hua2* single mutants. While I used

Figure 3.29. Two-hybrid assay for phosphorylation-dependent binding of HUA2 to GAL4 AD-CTD fusions. PJ69-4A cells were co-transformed with constructs encoding the indicated fusion proteins. Transformants were grown on selection plates lacking leucine, tryptophan and histidine (-LUH).

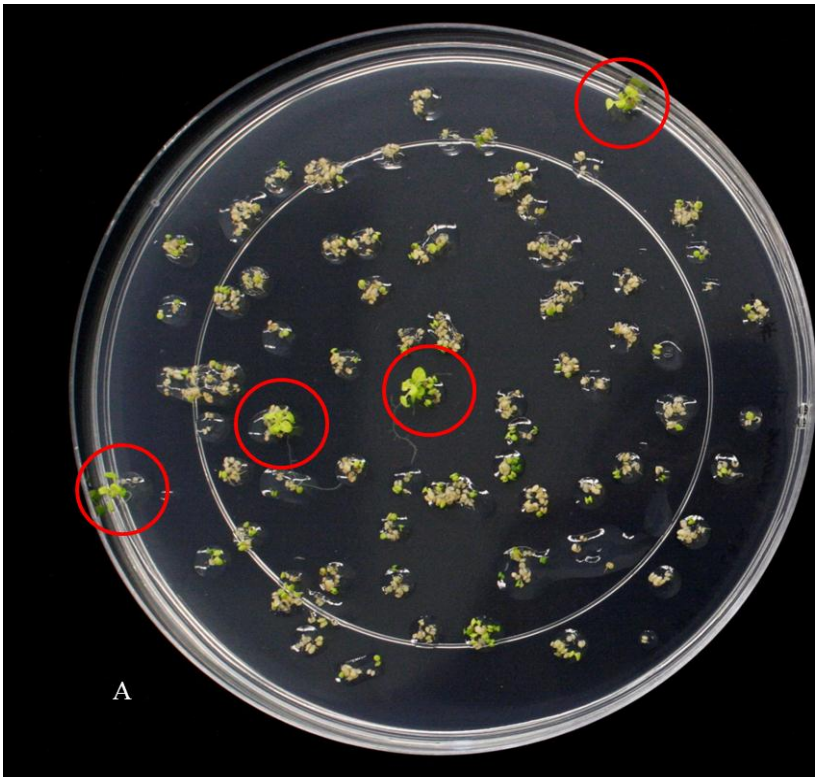
GAL4DB/AD fusions



FRI hua2-3 lines for scoring flowering phenotype, I used *hua2-7/hulk1* double mutants to score for embryo defect phenotype. I was curious to determine if overexpression of wild-type *HUA2* or *HUA2ΔRPR* mutant could rescue the early flowering observed in *FRI-hua2-3*, and the embryo defect observed in *hua2-7/hulk1* double mutants. To determine this, I overexpressed *HUA2* or *HUA2ΔRPR* under the control of the CaMV35S promoter in *FRI-hua2-3* lines and *hua2-7/hulk1* double mutants. Empty vector controls were also tested in the same lines. Positive T1 transgenic lines from both batches were selected on MS plates supplemented with hygromycin (Figure 3.30). Two weeks after germination, 36 T1 transgenic plants each transformed with, *35S:HUA2, 35S:HUA2ΔRPR* and the vector control were transplanted to soil. For the analysis of flowering time for *FRI hua2-3* genotype, the total number of basal rosette leaves for 18 out of 36 independent T1 plants was recorded and analyzed. Early flowering plants display lower number of basal rosette leaves relative to late flowering strain. The mean number of rosette leaves at bolting was significantly lower in both vector control and *35S:HUA2ΔRPR* lines relative to *35S:HUA2* lines ($p<0.0001$). There was no significance difference between the vector control and *35S:HUA2ΔRPR* lines ($p=0.4763$) (Figure 3.31 and 3.32). This data suggests that overexpression of *35S:HUA2* is capable of complementing late flowering in FRI *hua2-3* thereby reconstituting the late flowering phenotype observed in FRI Col-0. However, overexpression of *35S:HUA2ΔRPR* could not promote late flowering suggesting that the DSI motif of the RPR domain is crucial for HUA2 function with respect to flowering time in *FRI hua2-3* background.

For the analysis of embryo defect the embryonic development of the T1 plants from *hua2-7/hulk1* double mutant background was observed. Out of the 36 independent T1

Figure 3.30. Selection of *hua2-7/hulk1* transformants. (A) The image shows seedlings from *hua2-7/hulk1* T1 seeds obtained from plants subjected to floral dip transformation using *Agrobacterium tumefaciens* strain GV3101 harboring the binary plasmid. Seeds were plated on 1% agar containing MS medium (1% sucrose) and hygromycin B at a concentration of 30 $\mu g.ml^{-1}$. Following a stratification period of two days at 4°C in the dark, seedlings were incubated for 6 h at 22°C in white light, followed by incubation at 22°C in the dark for two days and continuous white light for a further 24 h. **(B)** Hygromycin B-resistant seedlings exhibit long hypocotyls whereas the non-transformants have short hypocotyls.


A

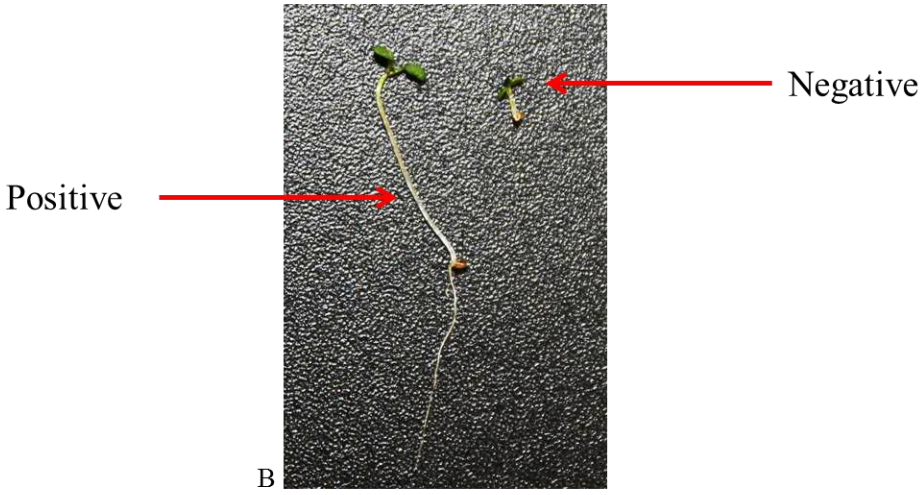
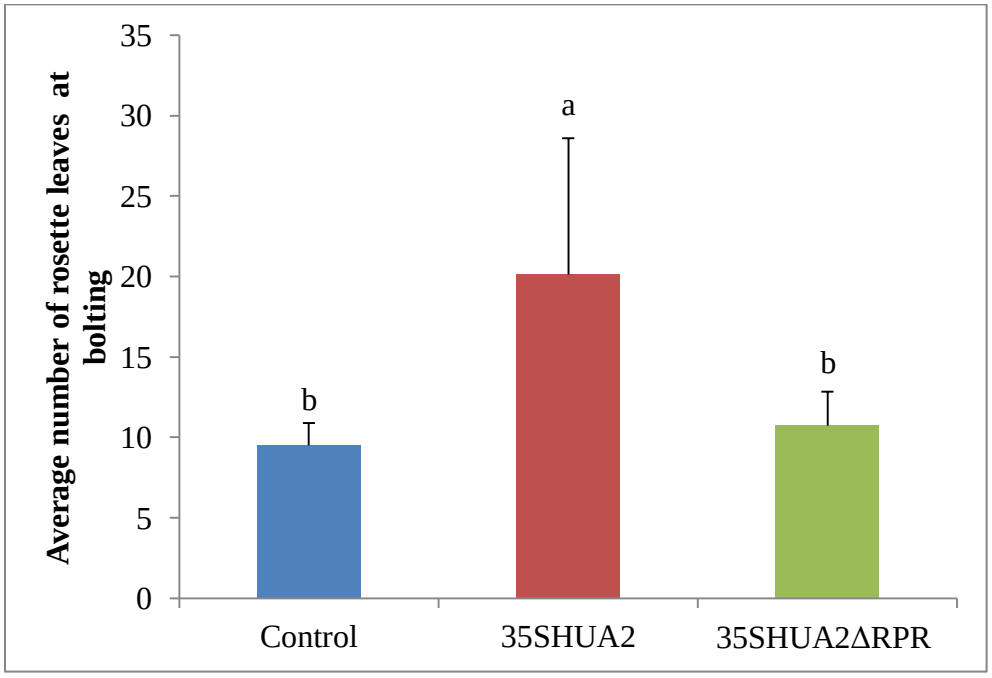
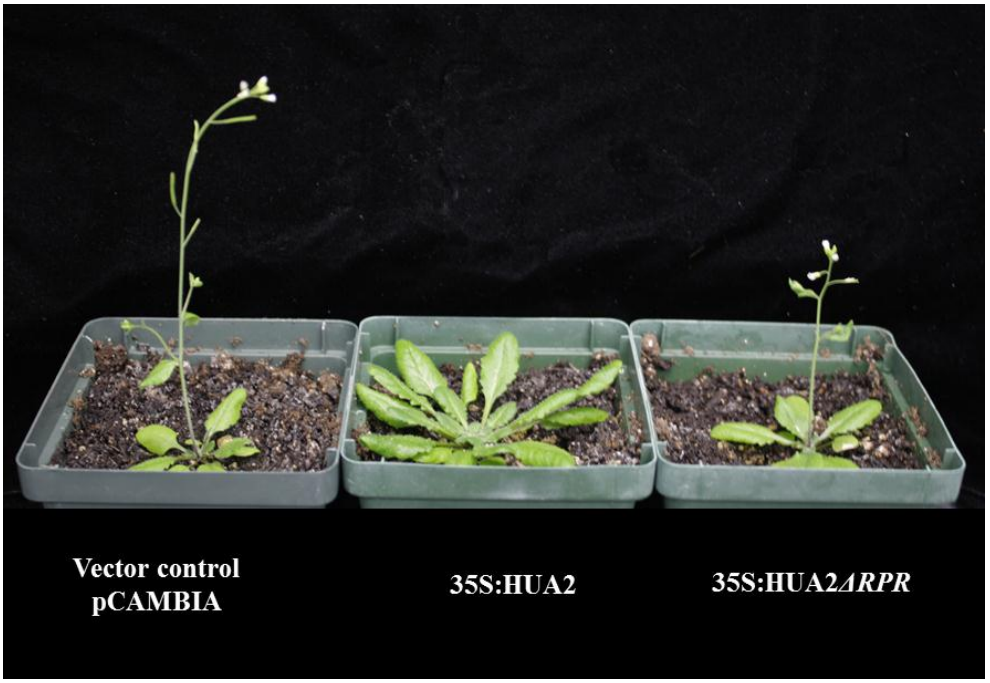

Negative
Positive
B

Figure 3.31. Flowering time of *Arabidopsis FRI-hua2-3* plants transformed with with pCAMBIA(vector control), *35S:HUA2* and *35S:HUA2ΔRPR*. Flowering time is measured by counting the total number of basal rosette leaves at the time of bolting. Bars represent means and error bars represent the standard deviation, n=18, $p<0.0001$ between 35SHUA2 and vector control/35SHUA2ΔRPR; $p=0.4763$ between vector control and 35SHUA2ΔRPR. Bars with different letters are statistically significantly different.

Figure 3.32. 30-day-old *Arabidopsis FRI-hua2-3* plants transformed with *35S:HUA2, 35S:HUA2ΔRPR* and pCAMBIA(vector control).


Average number of rosette leaves at bolting
35
30
25
20
15
10
5
0
a
b
b
Control
35SHUA2
35SHUA2ΔRPR


Vector control pCAMBIA
35S:HUA2
35S:HUA2ΔRPR

transgenic plants from *hua2-7/hulk1* double mutant background, 10 independent plants from each line were used for the analysis of embryo phenotype. The number of non-defective embryos was significantly higher in *35S*:*HUA2* lines compared with the vector control and *35S*:*HUA2ΔRPR* lines. In contrast, there was no significant difference in non-defective and defective embryos between vector control and *35SHUA2ΔRPR* lines (refer to Table 3.10 for p values). These data suggests that the overexpression of *35S:HUA2* rescues the embryo defect phenotype in *hua2-7/hulk1*double mutant backgrounds. On the other hand, overexpression of *35S*:*HUA2ΔRPR* was unable to rescue the embryo defective phenotype in *hua2-7/hulk1* background suggesting that the DSI motif within the RPR domain of HUA2 is required to rescue the embryo defect phenotype (Figure 3.33 and 3.34).

3.12 Significance of PWWP domain of HUA2 using genetic complementation test

I next determined the significance of the PWWP domain within HUA2. To this end, I performed a complementation test using *FRIhua2-3* genotype. In order to check the importance of the PWWP domain within HUA2, I overexpressed *HUA2* or *HUA2ΔPWWP* in *FRIhua2-3* to determine if it could reconstitute the late flowering phenotype observed in *FRI*-Col. *HUA2* and *HUA2ΔPWWP* were overexpressed under the control of the CaMV35S promoter in *FRIhua2-3* plants. As a control empty vector pEarlygate 100 was also transformed into *FRI hua2-3* plants. The positive T1 transgenic lines were selected on soil using BASTA selection (Figure 3.35). Two weeks after germination, 36 T1 transgenic plants for each construct, *35S*:*HUA2*, *35S*:*HUA2ΔPWWP* and the vector control were transplanted individually on to new pots with soil. For the analysis of the flowering time phenotype, the total number of basal rosette leaves of all.

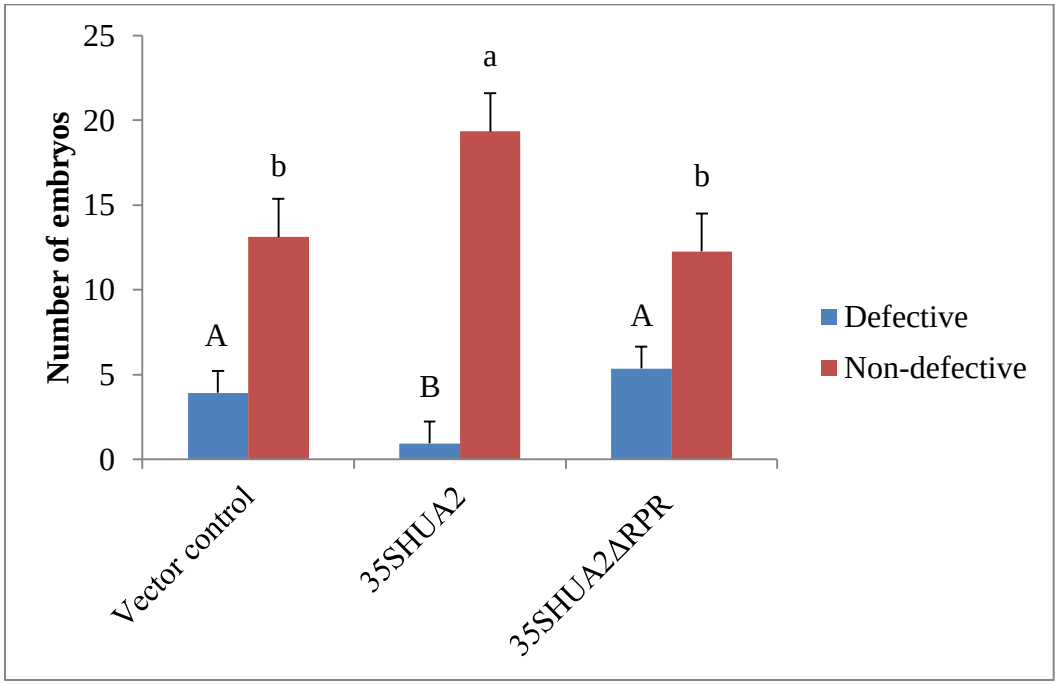
Table 3.10. Results from the one-way ANOVAs performed for the mean number of wild-type and defective embryos.

Number of defective embryos ANOVA	p-value	r^2	Alpha
35S:HUA2 vs. *35S:ΔRPR* vs. Vector Control	0.0003	0.4562	0.05
Post-hoc test			
35S:HUA2 vs. *35S:ΔRPR*	<0.0001		0.0167[1]
35S:HUA2 vs. Vector Control	0.004		0.0167
35S:ΔRPR vs. Vector Control	0.1399		0.0167
Number of wild-type embryos ANOVA			
35S:HUA2 vs. *35S:ΔRPR* vs. Vector Control	0.0036	0.3414	0.05
Post-hoc test			
35S:HUA2 vs. *35S:ΔRPR*	0.002		0.0167
35S:HUA2 vs. Vector Control	0.0056		0.0167
35S:ΔRPR vs. Vector Control	0.6794		0.0167

[1]Alpha adjusted for multiple comparisons using Bonferroni correction.

Figure 3.33. Embryo defects in *Arabidopsis hua2-7/hulk1* plants transformed with pCAMBIA(vector control), *35S:HUA2* and *35S:HUA2ΔRPR*. Phenotypes are characterized as the number of defective and wild-type embryos. Bars represent means and error bars represent the standard deviation, n=10. Bars with different letters are statistically significantly different.

Figure 3.34. Representative siliques from wild-type Col-0, *hua2-7/hulk1*plants transformed with pCAMBIA(vector control), *35S:HUA2* and *35S:HUA2ΔRPR*. Siliques were split open by hand to reveal the developing seeds. Red asterisks indicate defective seeds.


Number of embryos
25
20
15
10
5
0
A
b
B
a
A
b
Vector control
35SHUA2
35SHUA2ΔRPR
Defective
Non-defective

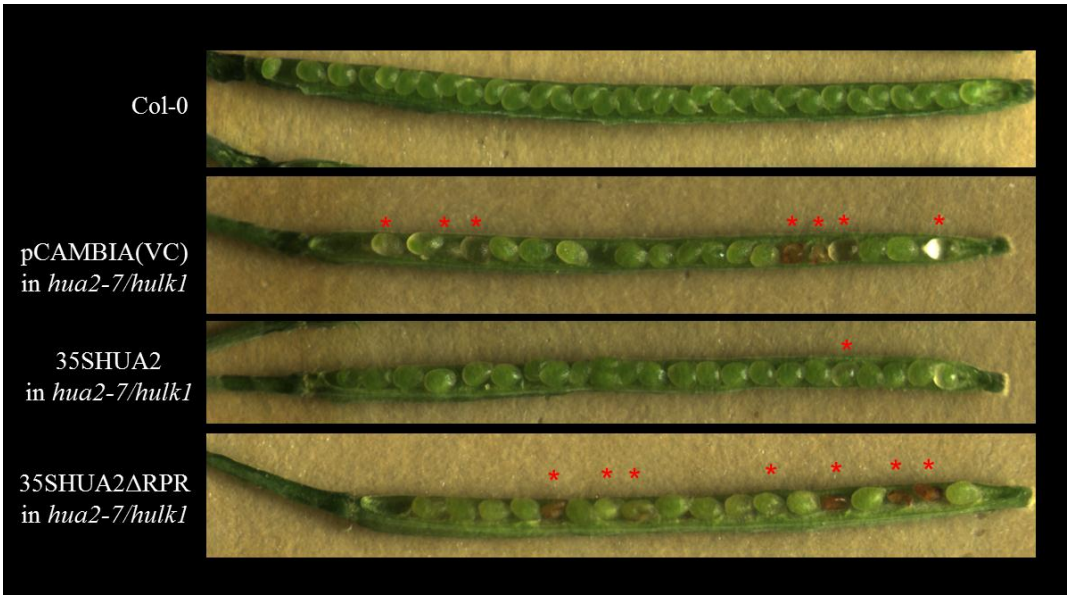
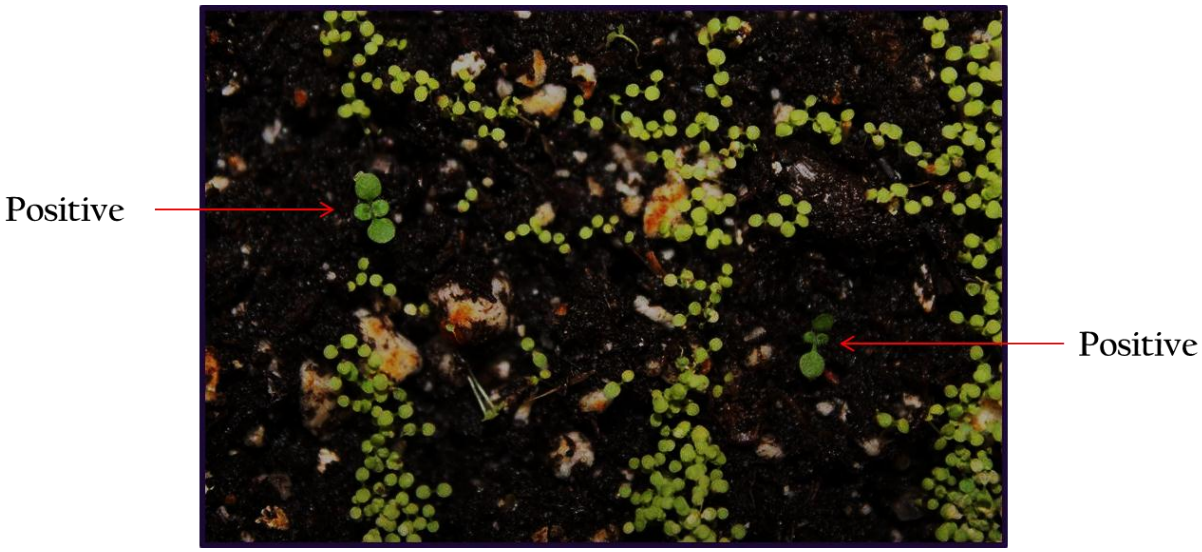

Col-0
pCAMBIA(VC)
in hua2-7/hulk1
35SHUA2
in hua2-7/hulk1
35SHUA2ΔRPR
in hua2-7/hulk1

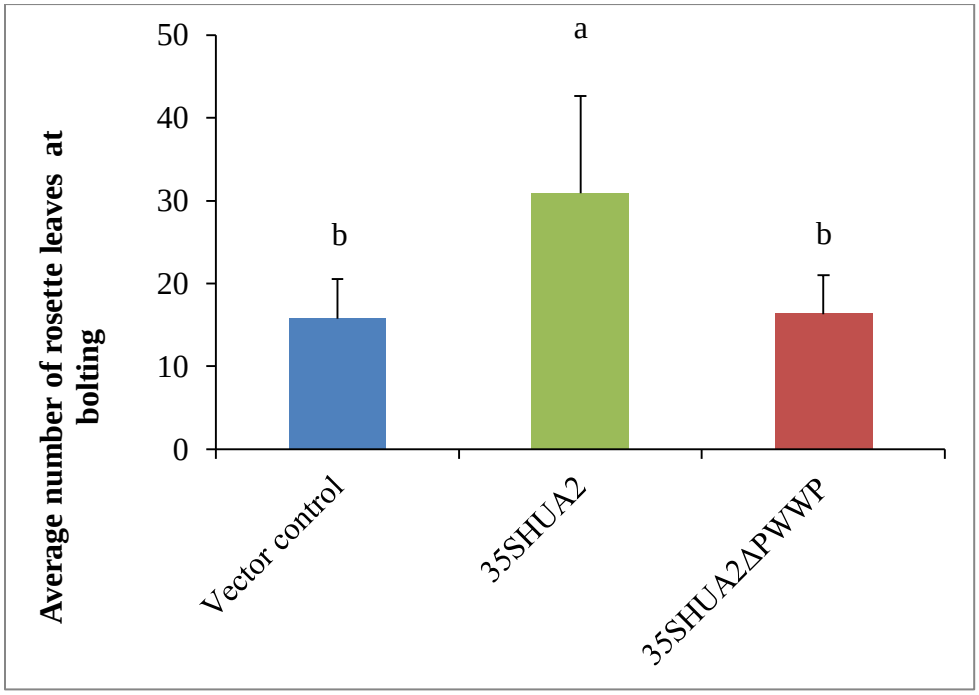
Figure 3.35. Selection of *FRI hua2-3* transformants. The image shows seedlings from *FRI hua2-3* T1 seeds obtained from plants subjected floral dip transformation using *Agrobacterium tumefaciens* strain GV3101 harboring the binary plasmid. Seeds were sprinkled on pots containing wet soil, followed by stratification for of two days at 4°C in the dark, then moved back to growth chamber. 10 days after germination BASTA 120 mg/L was sprayed on seedlings every alternate day. Red arrows indicate positive transformants.

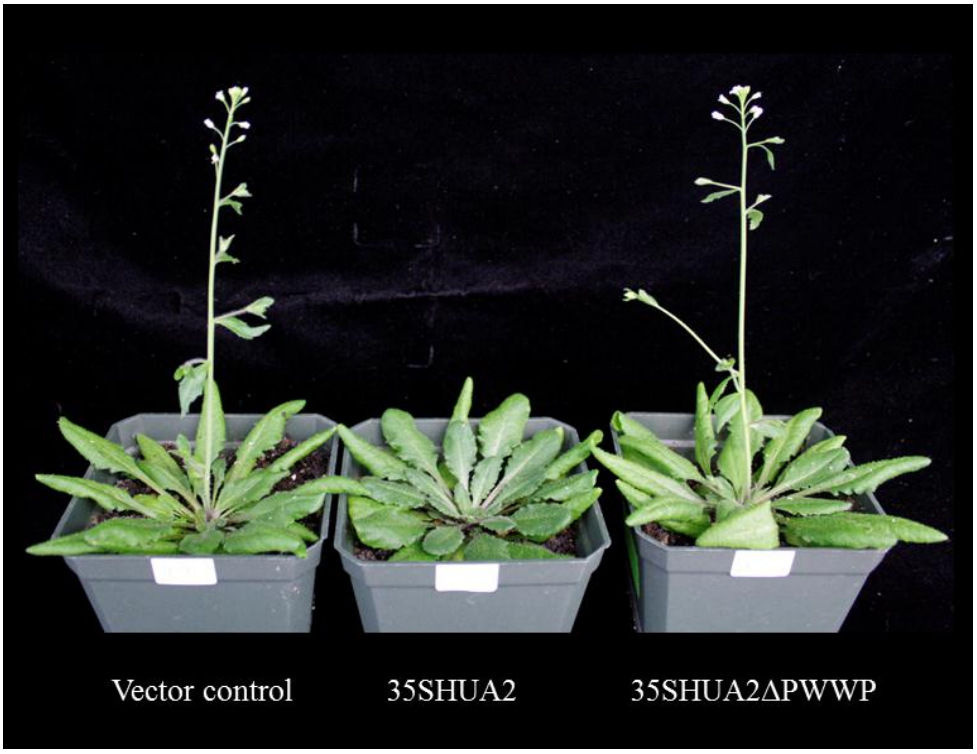

Positive
Positive

36 T1 plants for each line was recorded and analysed. The mean number of rosette leaves at bolting was significantly lower in both the vector control and the *35S:HUA2ΔPWWP* lines relative to *35S:HUA2* lines ($p<0.0001$). There was no significant difference between the vector control and the *35SHUA2ΔPWWP* lines ($p =0.5709$) (Figure 3.36 and 3.37). This data suggests that overexpression of *35S:HUA2* is capable of inducing late flowering in FRI *hua2-3* thereby reconstituting the late flowering phenotype observed in FRI Col-0. However, overexpression of *35S:HUA2 ΔPWWP* could not promote the same level of late flowering suggesting the PWWP domain is crucial for HUA2 function with respect to flowering time.

Figure 3.36. Flowering time of *Arabidopsis FRIhua2-3* plants transformed with pEarlygate100 (vector control), *35SHUA2* and *35SHUA2ΔPWWP*. Flowering time is determined by measuring the total number of basal rosette leaves at the time of bolting. Bars represent means and error bars represent the standard deviation, n=36, $p<0.0001$ between vector control/35SHUA2ΔPWWP and 35SHUA2; $p=0.7636$ between vector control and 35SHUA2ΔPWWP. Bars with different letters are statistically significantly different.

Figure 3.37. 30-day-old *Arabidopsis FRIhua2-3* plants transformed with *35S:HUA2*, *35S:HUA2ΔPWWP* and pEarlygate100 (vector control).


Average number of rosette leaves at bolting
50
40
30
20
10
0
a
b
b
Vector control
35SHUA2
35SHUA2ΔPWWP


Vector control
35SHUA2
35SHUA2ΔPWWP

Chapter 4: Discussion

4.1 The molecular function of HUA2: a brief summary of early genetic evidence.

HUA2 is a plant specific gene that was first identified along with *HUA1* in a genetic screen aimed at identifying partially redundant genes that function in the *AG* floral organ identity pathway in *Arabidopsis*(Chen and Meyerowitz, 1999; Li et al., 2001). While it was clear that HUA1 and HUA2 play an important role in floral development it should be noted that *hua1 hua2* double mutant display a variety of pleiotropic phenotypes and that both *HUA1* and *HUA2* exhibit broad expression patterns. Thus, these data suggest that the functions of *HUA1* and *HUA2* are not necessarily restricted to floral organ identity and /or development (Chen and Meyerowitz, 1999; Li et al., 2001; Challa, unpublished).

HUA1 is a nuclear protein that exhibits similarity to CCCH zinc-finger proteins and is capable of binding to RNA and single-stranded DNA, but not to double stranded DNA *in vitro*. The molecular function of HUA2, on the other hand, remains more obscure. Genetic evidence has shown that *HUA2* participates in a variety of processes including vegetative development and floral transition through its interactions with the *FRI* and *FLC* genes (Chen and Meyerowitz, 1999; Poduska et al., 2003; Doyle et al., 2005; Wang et al., 2007). It has also been shown to participate in floral specification through the regulation of *AG* expression (Chen and Meyerowitz, 1999). Furthermore, previous work from our laboratory has shown that the effects of *HUA2* on *FLC* and *AG* can be uncoupled. The current project has focused on further characterizing the molecular function of HUA2 (and other members of the HULK gene family) and implicates the HULK protein family to play an important role in the processing of primary transcripts in the nucleus.

4.2 The role of *HUA2* in regulating *FLC*

Previous work from our laboratory and other groups has established that HUA2 regulates *FLC* expression (Doyle et al., 2005; Wang et al., 2007). However, the mechanism of this regulation is unknown and could be either direct or indirect. In the following paragraphs I briefly summarize what is known regarding *FLC* expression and then propose and discuss potential models of HUA2 dependent *FLC* regulation

4.2.1 Regulation of *FLC*

Flowering time is an important aspect of the plant life cycle with respect to adaption and reproductive success. The *FLC* gene is a central component regulating the switch to reproductive development in *Arabidopsis*(Michaels and Amasino, 1999). It encodes a MADS box transcription factor that prevents flowering by repressing transcription of the down-stream flowering pathway integrator genes (Michaels and Amasino, 1999; Sheldon, 1999). The repressor activity is quantitative and there is a general correlation between levels of *FLC* expression and flowering time in *Arabidopsis* ecotypes (Rouse et al., 2002). The expression of *FLC* is regulated through several pathways by diverse mechanisms, including FRIc/AtPAF1c and/or lncRNAs mediated transcriptional regulation of the *FLC* locus.

4.2.2 FRIc/PAF1 mediated regulation of *FLC*

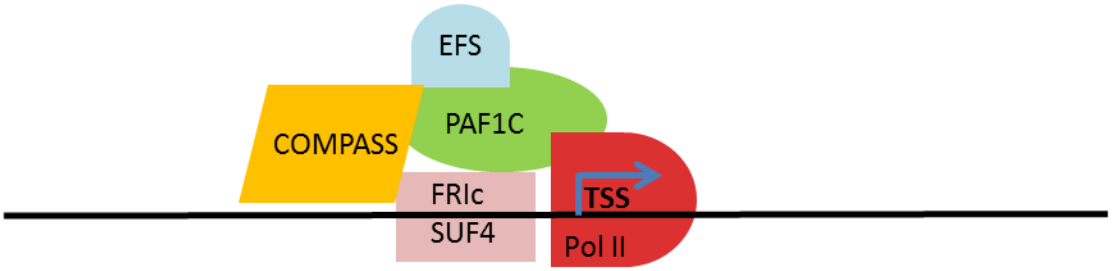
FLC transcription is negatively regulated by components of the vernalization and autonomous pathways and is positively regulated by *FRI*(Michaels and Amasino, 1999; Sheldon, 1999). Like *FLC, FRI* is a major determinant of the natural variation in flowering time and vernalization responses observed among *Arabidopsis* accessions. It

encodes a protein with two coiled-coil motifs and is required for high levels of *FLC* transcriptional activation (Johanson et al., 2000). Recent studies have shown that FRI together with FRL2, FES1, SUF4 and FLX4 form a putative transcriptional activator complex called FRIc and that SUF4 of the FRIc complex recognizes a *cis*-element in the *FLC* proximal promoter (Choi et al., 2011). Furthermore, FRIc directly associates with EFS, a multi-catalytic histone methyltransferase which is involved in trimethylation of Lysine 4 on Histone 3 (H3K4me3) on *FLC* chromatin (Ko et al., 2010b).

H3K4me3 is found at actively transcribed promoters, particularly downstream of the transcription start site, and is correlated with transcription activation. It has also been suggested that several chromatin modifiers that bring about histone acetylation, H3K4me3, H3K36me2 and me3 are associated with *FRI*-dependent *FLC* activation (reviewed in He, 2012). FRIc, like the *Arabidopsis* PAF1 (PAF1 for RNA Polymerase II Associated Factor 1) complex (AtPAF1c), is thought to recruit the COMPASS complex, which initiates H3K4 methylation activity at *FLC* chromatin (Jiang et al., 2009).

The *Arabidopsis* PAF1 complex is composed of six subunits, functional disruption of any one of these subunits leads to reduction in H3K4me3 at *FLC* chromatin, supressing the expression of *FLC* even in *FRI* background (He et al., 2004; Xu et al., 2008). Since both FRIc and AtPAF1 are associated with COMPASS, there is likely a functional dependence between the two complexes (reviewd in He, 2012). Figure 4.1 shows a model through which *FLC* transcript levels are regulated by FRIc/PAF1 complex at the proximal *FLC* promoter region.

Figure 4.1 Model showing the FRIc/PAF1 dependent *FLC* activation. Model of FRIc/PAF1-mediated recruitment or enrichment of chromatin modifiers at the *FLC* locus at the transcription start site (TSS).The FRIc, through SUF4 subunit, binds to the proximal FLC promoter, further increasing the binding of COMPASS for histone H3 lysine-4 trimethylation (H3K4me3) to the region around the TSS. In addition, PAF1c in association with RNA Pol II enriches COMPASS and EFS by directly associating with FRIc at the *FLC* locus. Model adapted and modified from (He, 2012).



FLC Chromatin

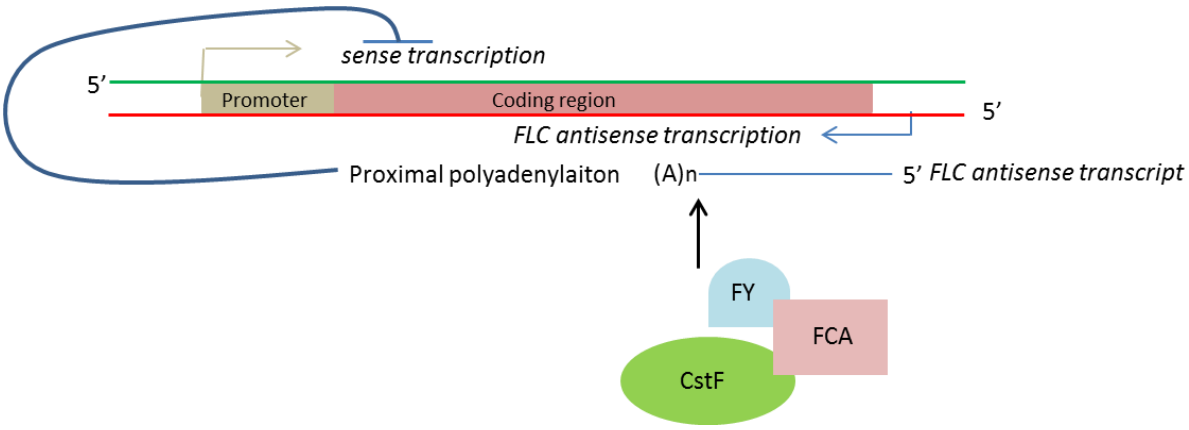
4.2.3 lncRNAs mediated regulation of *FLC*

Recent studies have also shown that the expression of *FLC* is negatively regulated through long noncoding lncRNAs that originate from *FLC* antisense transcripts (Liu et al., 2007). lncRNAs are regulated by the RNA binding proteins FCA and FPA and the RNA 3'-end processing factors FY, CstF64, and CstF77 (Liu et al., 2010; Hornyik et al., 2010). There are two different classes of antisense *FLC* transcript: Class I and Class II, that are produced as a result of alternative 3' polyadenylation, both of which are co-expressed at the *FLC* locus (Liu et al., 2010). Polyadenylation occurs at the proximal poly(A) site in Class I and at the distal site in Class II RNAs (Liu et al., 2010; Hornyik et al., 2010). It has been shown that the CstF-3'-end processing complex containing CstF67 and CstF77, in association with FPA, FCA and FY promotes the activity at the proximal poly(A) site, leading to the production of Class I antisense transcripts. While the production of Class I antisense transcripts affects the chromatin state at *FLC* locus and results in *FLC* suppression, production of Class II is associated with high H3K4me2 levels in the coding region of *FLC* which results high expression of *FLC*(Liu et al., 2010). Figure 4.2 shows a model of *FLC* regulation through lncRNAs.

4.3 Direct and indirect models of HUA2 dependent *FLC* gene regulation.

HUA2 may affect the regulation of *FLC* gene expression in either a direct or an indirect manner. Direct regulation may occur in one of two different ways: 1) HUA2 can directly associate with the *FLC* chromatin and together with FRIc/AtPAF1 complexes affect *FLC* expression; or, 2) HUA2 can regulate the expression of *FLC* antisense transcripts acting

Figure 4.2 Model for chromatin silencing of *FLC* through regulation of its antisense transcript. The CstF-mediated 3'-end processing of *FLC* antisense transcripts at the proximal poly(A) site, promoted by FCA-FY. Model adapted and modified from (He, 2012).


sense transcription
5′
Promoter
Coding region
5′
FLC antisense transcription
Proximal polyadenylaiton
(A)n
5′ FLC antisense transcript
FY
FCA
CstF

antagonistically to FCA-FY/CstF complexes and promote utilization of a distal 3'end of the lncRNAs which is associated with an increase in H3Kme3 levels in the coding region of *FLC*, thereby increasing *FLC* levels. In addition, since these mechanisms are not mutually exclusive, there is a possibility that HUA2 is involved in both of the above processes. Alternatively, HUA2 may be acting indirectly by regulating the expression of genes in the flowering pathway, which in turn regulate *FLC* expression. The putative mechanisms through which *HUA2* regulates *FLC* expression are further discussed below.

4.3.1 Role of HUA2 in FRIc/AtPAF1 regulation of *FLC*

Yeast two-hybrid screens performed in our laboratory and by other groups have provided no evidence of HUA2 interaction with the components of FRI or AtPAF1 complexes. This suggests that HUA2 is not part of these complexes. However, our laboratory has previously shown that HUA2, along with components of AtPAF1 complex, are required for H3K4me3 at the *FLC* locus and that loss of *HUA2* leads to reduction of *FLC* primary transcripts in plants containing *FRI*. Presence of *FRI* in *HUA2* backgrounds elevates H3K4me3 levels at the *FLC* chromatin (Grbic, unpublished). This suggests that HUA2 may associate with the members of FRIc and/or PAF1c to participate in the recruitment and/or enrichment of H3K4 methyltransferase at the *FLC* chromatin.

If there is a direct interaction between HUA2 and *FLC* chromatin, it could be mediated via the PWWP domain of HUA2. The PWWP domain is a member of the Tudor domain 'Royal Family' which is highly conserved among the eukaryotic kingdoms (Qiu et al., 2002). The *Arabidopsis* genome contains around 16 genes that encode proteins containing the PWWP domain (Alvarez-Venegas and Avramova, 2012). Studies done so

far on different yeast and human proteins having PWWP domains suggest that they interact with histone tails (Ge et al., 2004; Wang et al., 2009; Vezzoli et al., 2010).

In this study, I showed that the PWWP domain within HUA2 is associated with control of flowering time. Using site directed mutagenesis and genetic complementation tests I demonstrated that the disruption of the PWWP domain within HUA2 results in HUA2 losing the ability to promote a late flowering phenotype in *FRI* Col-0. *FRI hua2-3* is an early flowering genotype relative to *FRI*-Col-0 (Michaels and Amasino, 2001). While the overexpression of *HUA2ΔPWWP* supresses late flowering, overexpression of *HUA2* with a functional PWWP domain promotes a late flowering phenotype in *FRI* Col-0(Figure 3.36 and 3.37). This indicates that the PWWP domain within HUA2 is required for the increase in *FLC* levels leading to the late flowering phenotype.

I thus propose that HUA2 is associated with the members of FRIc and/or PAF1c and participates in the recruitment and/or enrichment of H3K4 methyltransferase by directly interacting with *FLC* chromatin. I speculate that this interaction occurs through the PWWP domain of HUA2. It has been previously shown that mutations within the PWWP domain of Dnmt3a abolishes its association with heterochromatin (Chen et al., 2004). Therefore, it is possible that mutation in PWWP of HUA2 may inhibit its association with the *FLC* chromatin. This might further restrict chromatin modifiers like the histone methyltransferase of the FRIc/AtPAf1 complexes from binding to the *FLC* locus, resulting in reduced levels of H3K4me3. Thus, if HUA2 affects *FLC* expression directly through the interaction with the FRIc/AtPAf1 complexes, then the expectation is that HUA2 with a functional PWWP domain will associate with the *FLC* chromatin while HUA2ΔPWWP will not.

4.3.2 Role of HUA2 in lncRNA regulation of *FLC*

Using site directed mutagenesis and genetic complementation assays I have shown that disruption of the RPR domain within HUA2 also causes HUA2 to lose its ability to promote a late flowering phenotype in *FRI* Col-0. While the overexpression of *HUA2ΔRPR* domain supresses late flowering, overexpression of *HUA2* with a functional RPR domain promotes a late flowering phenotype in *FRI* Col-0(Figures 3.31-3.32). This indicates that, similar to the PWWP domain, the RPR domain is important for the control of flowering time, and further suggests a possible pathway through which HUA2 regulates the expression of *FLC* through the RPR domain.

RPR domain containing proteins are known to function in nuclear pre-mRNA regulation (Amrani et al., 1997; Steinmetz and Brow, 1998; Yuryev et al., 1996). RPR domain containing yeast proteins such as Pcf11, Nrd1 and certain SR proteins such as RA4 and RA8 have been shown to interact with the CTD of RNA pol II (Yuryev et al., 1996; Steinmetz et al., 2001; Arigo et al., 2006). This raises the possibility that HUA2 might also be associated with RNA pol II complex. In this study, I present data demonstrating that HUA2 does indeed interact with the CTD of RNA Pol II in yeast. Furthermore, I show that this interaction is phosphorylation-dependent. This suggested for the first time that HUA2 is a phospho-CTD associated protein (PCAP) in *Arabidopsis*. The implications of these observations are discussed further below.

4.3.3 HUA2 interacts with the Arabidopsis RNA Pol II CTD

Through yeast two-hybrid assays, I show that full length HUA2 is capable of interacting with the *Arabidopsis* RNA pol II CTD (AtCTD) in yeast (Figure 3.24 and 3.27). Interestingly; however, interaction was not observed between CT-HUA2 and AtCTD. This indicates the interaction between HUA2 and AtCTD requires protein domains upstream of CT-HUA2 sequences (such as the RPR domain). The yeast cleavage and polyadenylation factor Pcf11 has been shown to require the highly conserved DSI motif within its RPR domain for interacting with CTD (Sadowski et al., 2003).

Unexpectedly, I found that the DSI motif within the RPR domain of HUA2 was not essential for its interaction with AtCTD in yeast. Both HUA2 and HUA2ΔRPR (with a disrupted DSI motif) both interacted with the AtCTD in yeast (Figure 3.24 and 3.27). However, the mutation in the DSI motif abolished the ability of HUA2 to delay flowering in plants when expressed in *hua2* mutant background. This indicates that an intact DSI motif is required for HUA2 function. At present, it is unclear why there is a discrepancy between yeast and plant data relative to the importance of the DSI motif. One possibility is that the RPR domain in the mutated version of HUA2-AAA did not fold properly in our experimental conditions in yeast and showed interaction with AtCTD.

4.3.4 Phosphorylation state of CTD affects its interaction with HUA2

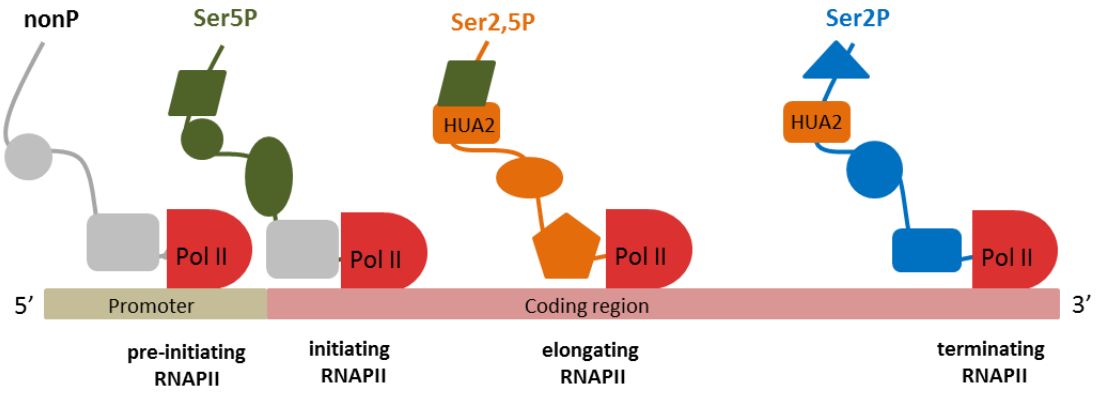
It is known that the phosphorylation states of Ser-2 and/or Ser-5 residues within the CTD play an important role in the recruitment of transcriptional components to the CTD that promote RNA synthesis and processing (Buratowski, 2005; Phatnani and Greenleaf, 2006; Saunders et al., 2006). Using CTD variants that differed in their ability to be

phosphorylated in a yeast two-hybrid assay (Ursic et al., 2008) I showed that HUA2 binds to CTD that is phosphorylated on Ser-2 residues (Figure 3.29). This suggests phosphorylation of Ser-2 in the heptad of CTD is necessary for HUA2 to interact with CTD.

The phosphorylation state of the AtCTD changes as RNA polymerase II progresses through the transcription cycle. In general, promoter binding and transcriptional initiation are carried out by non-phosphorylated CTD, while elongation of mRNA and its processing associate with hyper-phosphorylated CTD (Dahmus, 1994). The presence of multiple phosphorylation states at different stages of transcription allows the CTD to regulate its binding partners and, consequently, facilitates the timely recruitment of transcriptional components to the CTD for RNA maturation and/or processing. The CTD acts as a binding platform for a wide range of transcription factors, and the binding of these factors to RNA Pol II links them to different transcriptional stages (Phatnani and Greenleaf, 2006).

HUA2 shows a binding specificity for heptad repeats carrying Ser-2 phosphorylation; therefore, HUA2 may participate in the elongation, mRNA processing, and termination stages of transcription (Figure 4.3). It is interesting that the binding specificity of HUA2 for repeats carrying Ser-2 phosphorylation is similar to the RPR domain containing yeast protein Pcf11, which is a 3'-end poly-A/cleavage factor.HUA2 is also involved in 3'-end processing and splicing of *AG* and *FCA* (Chen and Meyerowitz, 1999; Uday, unpublished). Therefore, it is possible that HUA2 may also act as a 3'-end processing factor. Further biochemical experiments are necessary to determine if HUA2 does indeed perform this function directly.

Figure 4.3 CTD phosphorylation patterns decide which factors associate with RNA Pol II, different phosphorylation states are associated with different stages of transcription. RNA Pol II is in red and present at four positions along the gene. The wavy line represents the CTD and its color indicates different phosphorylation states: gray indicates non-phosphorylated repeats; green indicates Ser5P repeats; orange indicates doubly phosphorylated Ser2,5P repeats; and blue indicates Ser2P repeats. Proteins bound to a type of repeat are represented in the same color as the repeat suggesting it requires that particular phosphorylation state to bind. Figure adapted and modified from (Phatnani and Greenleaf, 2006).


nonP
Ser5P
Ser2,5P
Ser2P
HUA2
HUA2
Pol II
Pol II
Pol II
Pol II
5’
Promoter
Coding region
3’
pre-initiating RNAPII
initiating RNAPII
elongating RNAPII
terminating RNAPII

4.3.5 The HUA2 proline-rich C-terminal region

Along with the PWWP and the RPR domain, the proline-rich C-terminal region of HUA2 is also associated with flowering time. Loss of function alleles *hua2-1* and *hua2-5* which lack the proline-rich C-terminal region of *HUA2*, confer early flowering accompanied by lower levels of *FLC* expression (Wang et al., 2007) suggesting the importance of the proline-rich C-terminal end of HUA2 with respect to flowering. Using yeast two-hybrid assays, I show that HUA2 interacts with pre-mRNA processing factors AtPRP40, FCA, RBP45 and UBP1 (Figure 3.1). In addition, using yeast-hybrid assay (Figure 3.4A and B) and *in vitro* GST-pull down assay (Figure 3.13) I further show the proline-rich C-terminal end of HUA2 is necessary for the protein-protein interaction.

AtPRP40 and FCA are both WW containing proteins categorized as PCAP proteins. AtPRP40 is a plant homolog of yeast splicing factor PRP40 which has been shown to interact with both phosphorylated and unphosphorylated CTD *in vitro* and localize within the nucleus (Kang et al., 2009). As mentioned earlier, FCA, a regulator of flowering time, controls 3'-end formation of specific transcripts (Quesada et al., 2003).

UBP1 and RBP45 are proteins containing the RNA recognition motif (RRM), previously identified from *Nicotiana plumbaginifolia*(Lambermon et al., 2000; Lorković, Wieczorek Kirk, Klahre, et al., 2000). RRM is the most common domain found in several eukaryotic proteins involved in RNA metabolism (Anantharaman, 2002). It is thought to function in post-transcriptional regulation of gene expression including RNA processing, which is involved in tissue-specific and/or developmental regulation in plants (Siomi and Dreyfuss, 1997; Albà and Pagès, 1998). RBP45 has triple RRMs and has been implicated in various stages of RNA metabolism, including stability and splicing (Lorković,

Wieczorek Kirk, Klahre, et al., 2000). AtRBP45 is the functional homolog of tobacco RBP45, and has a strong sequence similarity and expression pattern to its tobacco counterpart (Peal et al., 2011). Similar to tobacco RBP45, AtRBP45 also associates with poly-A RNA (Lorković, Wieczorek Kirk, Klahre, et al., 2000; Schmidt et al., 2010). In addition to its role in RNA metabolism, it has also been shown to be involved in repression of translation in mature pollen grains in *Arabidopsis* and has been suggested to participate in several undefined steps of pre-mRNA maturation in plants (Lorković, Wieczorek Kirk, Klahre, et al., 2000; Park et al., 2006).

Like RBP45, UBP1 also has three RRMs and has been observed to enhance the splicing of suboptimal introns. The mRNA 3'-UTR region may also be associated with UBP1 and likely protects the mRNA from degradation and/or is required for pre-mRNA maturation. Similar to RBP45, UBP1 associates with the poly-A RNA *in vitro* and has been suggested to be an hnRNP-like protein (Lambermon et al., 2000). Several hnRNPs are associated with recognition of the U-rich sequences that facilitate spliceosomal association. Plant introns contain high proportions of UA or U rich sequences and the 3'-end of the introns are rich in uridine residues, particularly in dicots. These U rich sequences play an important role in the recognition of plant introns (Goodall and Filipowicz, 1989; Brown and Simpson, 1998). Mutation analysis has shown that uridines are essential for splice site selection (Merritt et al., 1997; Ko et al., 1998). The *Arabidopsis* homolog AtUBP1 has been shown to interact with two other plant specific proteins, UBA1 and UBA2. UBA1 and UBA2 are nuclear proteins that may belong to a complex which contributes to the stabilization of mRNA in the nucleus (Lambermon et al., 2002). Taken together, the interaction between HUA2 and AtPRP40, FCA, RBP45

and UBP1 provides further evidence that HUA2 is associated with splicing and/or 3'-end processing. I hypothesize that HUA2 is involved in the regulation of 3'-end formation of *FLC* antisense transcripts, similar to its involvement in the 3'-end processing of *FCA* and *AG* through the RPR domain and the proline rich C-terminal region. At this point, the exact mechanism through which HUA2 affects this process remains unknown.

4.3.6 Indirect models of HUA2 dependent *FLC* gene regulation.

I have shown that HUA2 interacts with pre-mRNA processing factors AtPRP40, FCA, RBP45 and UBP1 and that HUA2 affects splicing and 3'-end processing of *AG* and *FCA*(Chen and Meyerowitz, 1999; Sajja, unpublished). *AG* and *FCA* may not be the only genes whose expression is affected by HUA2. Given that HUA2 interacts with general splicing factors, it is possible that *HUA2* affects splicing and 3'-end processing of several other genes/transcripts that could regulate FLC expression. For example, HUA2 might indirectly affect *FLC* expression through regulation of *FCA* expression.

4.4 HULK gene family

HUA2 belongs to a small gene family. We have identified three other genes,*HUA2-LiKe (HULK1-3),* that share a high degree of sequence similarity and protein structure with *HUA2.* These genes have broad overlapping expression patterns during both vegetative and reproductive plant development and act redundantly. Except *hua2*, other single mutants do not display developmental abnormalities. Double and triple compound mutants progressively had more severe pleiotropic phenotypes that culminated in the lethality of the quadruple mutant, indicating that these genes act redundantly to support essential cellular functions (Challa, unpublished).

The *hua2-7/hulk1* line is the only double mutant line that displays pronounced phenotypic abnormalities. In addition to early flowering, developed siliques of *hua2/hulk1* plants have seeds with embryo defects, which are not observed in either of the single mutants (Challa, unpublished). In this study, I show that over expression of *HUA2* in *hua2/hulk1* double mutants can rescue the embryo defect phenotype, whereas overexpression of HUA2 with a disrupted DSI motif within the RPR domain fails to rescue the embryo defect observed in the *hua2/hulk1* double mutants (Figure 3.35 and 3.36). This supports the conjecture that *HUA2* is not only associated with the regulation of flowering time but is also associated with other developmental processes. At this point, we do not know if this is a direct or indirect effect. Just as with the flowering time phenotype, we do not know if disruption of the RPR domain within HUA2 affects its interaction with AtCTD and further affects AtCTD interaction with transcription machinery. Future experiments involving molecular characterization of the embryo defective phenotypes in *hua2-7/hulk1* mutants should help us understand the role of the *HUA2* in this process.

4.4.1 HULK proteins have conserved domain architecture and all localize in the nucleus and interact with same set of proteins as HUA2

The HULK proteins have high amino acid sequence similarity and share the same conserved domains and architecture with HUA2. All of the HULK proteins contain a PWWP domain, an RPR domain, a proline-rich region at the C-terminal end along with putative nuclear localization signals. Functional divergence of duplicated genes could be initiated by the divergence of gene expression domains within tissues or organs. In accordance with the known pleiotropic functions of *HUA2*, *HUA2* is expressed throughout the plant over the entire life cycle (Chen and Meyerowitz, 1998; Poduska et

al., 2003; Wang et al., 2007). *In situ* hybridization using gene specific probes for the remainder of the *HULK* family members on both vegetative and floral apices demonstrated that the expression domains of gene family members have not diverged. All four gene family members show similar broad expressions that were most concentrated at the vegetative and reproductive apices. In this study, I further investigated the subcellular localization of the HUA2 and HULK proteins by transient transgene expression and visualization of GFP/YFP tagged proteins in tobacco epidermal cells and showed that they all localize within the nucleus of the cell (Figure 3.19 and 3.20). In addition, I determined that a NLS resides at the amino terminus end of HUA2 (Figures 3.21 and 3.22).

Both HUA2 and the HULKs have proline-rich regions with PPLP repeats at the carboxyl end, suggesting that like HUA2, the HULKs might also participate in protein-protein interaction. In this study, I showed that the HULK proteins, like HUA2, participate in protein-protein interaction via their proline-rich regions. Using yeast two-hybrid assay, I show the C-terminal ends of HULK1 and HULK3 also interact with the same set of proteins as HUA2 (Figure 3.6). The high auto-activation of HULK2 made it unsuitable to be used in yeast hybrid assay, indicating the possibility of HULK2 containing an activation domain. Using GST pull-down assay, I was able to show that the C-terminal end of HULK2, like HULK1 and HULK3, also interacts with the same set of proteins as HUA2, Viz., FCA, AtPRP40, RBP45 and UBP1 (Figures 3.13, 3.14 and 3.15). This provides further evidence supporting their functional redundancy. In addition, yeast two-hybrid analysis showed that the HULK proteins, like HUA2, also interact with the CTD of RNA Pol II (Figure 3.28). The fact that HUA2 and the HULK proteins have the same

interacting partners within the nucleus suggested the possibility that HUA2 and HULKs may interact with each other. However, the yeast two-hybrid assays suggested this not to be so (Figure 3.17).

Although HUA2 and the HULK proteins interact with the same set of proteins, the strength of the interactions between them is not the same. Yeast two-hybrid analysis showed that the interaction between HUA2-PRP40 and HUA2-FCA is stronger than HUA2-RBP45 and HUA2-UBP1. HULK1 showed a stronger interaction preference with PRP40 and RBP45, whereas HULK3 shows a stronger binding preference with UBP1 and FCA. At this point, the reason why HUA2 and HULK proteins have differential preferences in interaction with the pre-mRNA processing factors is unknown. The strength of the interaction mainly depends on the nature of the interaction; they can be stable or transient. Since the HULK family members are associated with pleiotropic phenotypes, it is possible that the members of the HULK protein family transiently interact with different pre-mRNA processing factors, at different times and function dynamically in several regulatory pathways.

Chapter 5: Summary and future directions

The *HUA2* gene of *Arabidopsis thaliana* is known to participate in distinct developmental processes. For example, *HUA2* participates in the C-function of floral development through effects on the pre-mRNA processing of *AG* transcripts and in the timing of the transition to flowering of the apical and axillary meristems through the effects on the *FLC* transcript levels mediated by the *FRI* pathway (Poduska et al., 2003; Doyle et al., 2005; Wang et al., 2007). The pleiotropic nature of *HUA2*-regulated phenotypes is reflected in the ubiquitous expression profile of *HUA2* in plant tissues (Chen and Meyerowitz, 1999). However, the numerous roles of *HUA2* suggested by genetic interaction studies contrasts the subtle flowering time and plant stature phenotypes of laboratory-derived, loss of function *HUA2* alleles (Weigel, 2012).

The *Arabidopsis* genome has three additional genes with exhibiting high sequence similarity to *HUA2.* These three genes are referred to as *HULK* (*HUA2 LIKE*) genes and are referred to as the *HULK* gene family. *HULK* gene family members are distributed on chromosome 5 (*HUA2*, At5g23150 and *HULK1*, At5g08230), chromosome 2 (*HULK2*, At2g48160) and chromosome 3 (*HULK3*, At3g63070). Protein family members share the same conserved domain structures and significant amino acid sequence similarity. The *HULK* gene family members have similar expression domains (Challa, unpublished). In this study I provide evidence for their functional redundancy by showing that all members of HULK protein family localize within the nucleus where they interact with the same set of splicing factors (AtPRP40, FCA, RBP45 and UBP1). Furthermore, I show that the proline-rich region at the C-terminal ends of the HULK proteins is sufficient and required for interacting with these splicing factors.

I have also shown that HUA2 and the HULK proteins interact with AtCTD in yeast. Using HUA2 as the representative member, I show that HUA2 acts as phospho-CTD-associated proteins (PCAP), with binding specificity for CTD heptad repeats carrying Ser-2 phosphorylation. Ser-2 phosphorylation state is associated with transcription elongation, mRNA processing and 3'-end termination. Therefore, my data suggest that HUA2 is associated with transcription elongation, processing and 3'-end termination of the pre- mRNA. This is in agreement with data showing that HUA2 affects the 3'-end processing and splicing of *AG* and *FCA* (Chen and Meyerowitz, 1999; Uday, unpublished).

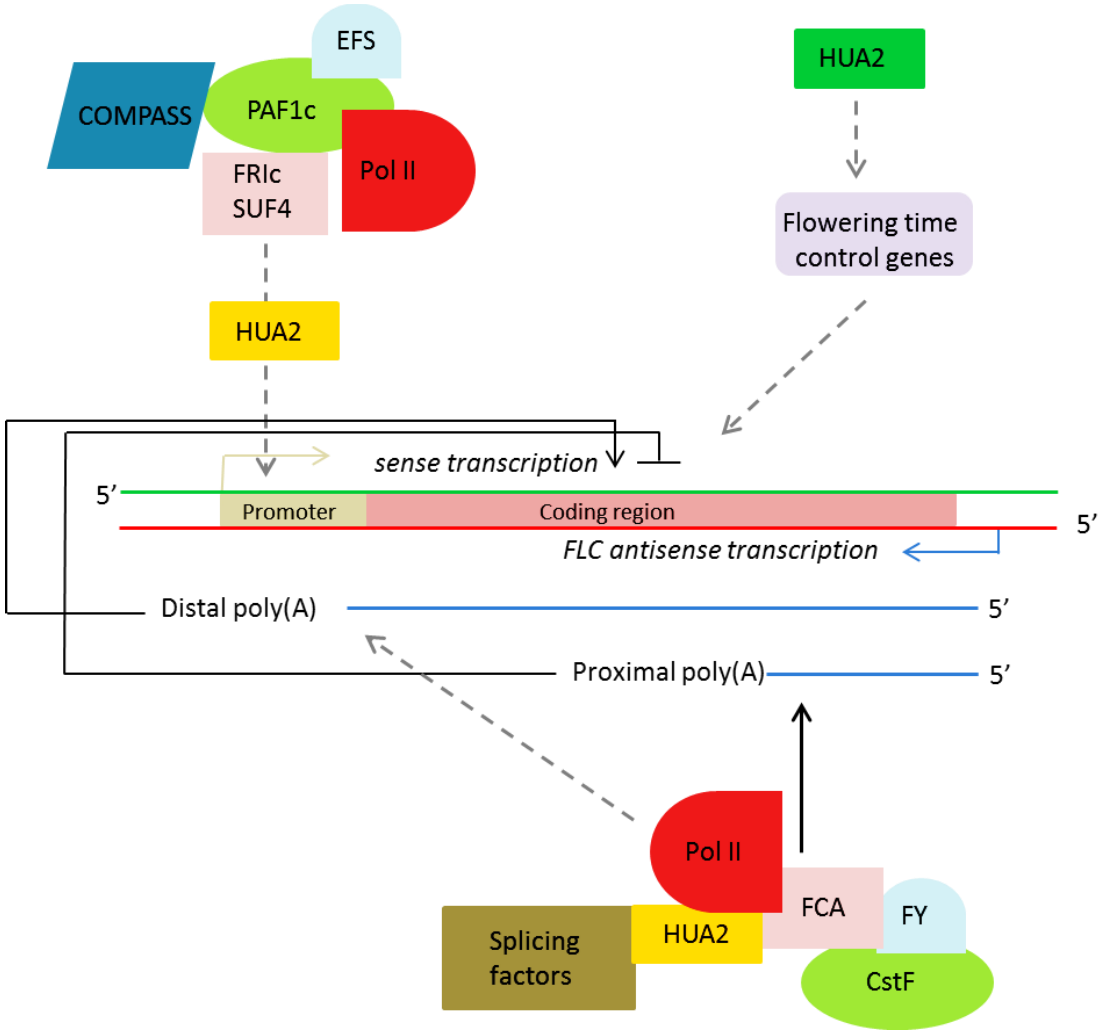
FLC encodes a MADS-box transcription factor which is a major repressor of flowering in *Arabidopsis*(Michaels and Amasino, 1999; Sheldon, 1999). It is one of the most extensively studied genes, because genetic variation at *FLC* and its activator *FRI* are responsible for the natural diversity in flowering among *Arabidopsis* ecotypes (Lee et al., 1993; Johanson et al., 2000; Michaels et al., 2003). *HUA2* has been shown to upregulate *FLC*; however, the mechanism of this genetic interaction is still unknown (Doyle et al., 2005; Wang et al., 2007). Along with *HUA2*, the *HULKs* also affect the steady state levels of *FLC* (Challa, unpublished).

Previous work on HUA2 suggested that the proline-rich C-terminal end of the protein is an important region for the ability of HUA2 to delay flowering (Wang et al., 2007). In this study, I have shown that the PWWP and the RPR domains are also important for the function of HUA2 to regulate the onset of flowering. The effect of HUA2 on regulating flowering is mediated through its effects on *FLC* expression, which can be direct or indirect. Based on the published data and results of my study, I propose a model (Figure

5.1) through which HUA2 might affect *FLC* regulation. HUA2 can affect *FLC* either directly through FRIc/AtPAF1c and lncRNAs pathways or indirectly by regulating the expression of genes in the flowering pathway, which in turn regulate *FLC* expression. The proposed model shows HUA2 to associate with the members of FRIc/AtPAF1c and participates in the recruitment of and/or enrichment of H3K4 methyltransferase by directly interacting with *FLC* chromatin via its PWWP domain. The second direct regulation is through the regulation of antisense transcripts of *FLC*. HUA2, in association with the CTD and splicing factors might affect the regulation of the two antisense transcripts, the distal polyadenylated and the proximal polyadenylated transcript. I further speculate that HUA2, acting antagonistically to FCA-FY/CstF complexes, promotes utilization of a distal 3'end of the lncRNAs which is associated with increasing *FLC* levels. It is possible that HUA2 regulates more than one discreet step in the regulation of the *FLC* expression through its association with different protein complexes. Given the redundant nature of the HULKs, it is possible for HULKs also participate to regulate *FLC*.

HUA2 has been shown to interact with other loci to control floral and vegetative development in addition to flowering time (Chen and Meyerowitz, 1999; Doyle et al., 2005; Wang et al., 2007). While single *HULK* mutant lines develop normal shoots, some compound mutants display morphological abnormalities (Challa, unpublished). The interaction of HUA2 and HULK proteins with the pre-mRNA processing factors and AtCTD may affect the regulation of many genes that are in involved in different developmental pathways. The fact that *hua2 hulk1 hulk2 hulk3* quadruple mutant

Figure 5.1 A putative model showing different mechanisms through which HUA2 regulates *FLC* expression. HUA2 can regulate the expression of FLC directly in two independent mechanisms through FRIc/AtPAF1c and /or through regulation of antisense transcripts in association FCA/FY CstF. HUA2 can indirectly regulate *FLC* expression by regulating other flowering time control genes. Direction interaction is showed with HUA2 in orange box and indirect regulation is shown with HUA2 in green box. Black arrows represent promotion and T-bars represent repression. Grey dotted arrows represent the regulatory relationship that was not directly analyzed in this work.


EFS
COMPASS
PAF1c
FRIc
SUF4
Pol II
HUA2
HUA2
Flowering time
control genes
sense transcription
5'
Promoter
Coding region
5'
FLC antisense transcription
Distal poly(A)
5'
Proximal poly(A)
5'
Pol II
FCA
FY
Splicing
factors
HUA2
CstF

(Challa, unpublished) are lethal is consistent with the possibility that HUA2 and HULK act as general pre-mRNA processing factors. The effect of HUA2 and HULKs on other genes may occur through mechanisms similar to those described for *FLC*. Future molecular and biochemical experiments need to be done to confirm these putative mechanisms. The following are some of the questions I would like to address in the future;

1) Does the disruption of the PWWP domain abolish or reduce H3K4me3 levels at *FLC* and further lead to its suppression? Are there other loci where the same phenomenon occurs?
2) Do the PWWP domains of HULKs perform similar functions as the PWWP domain of HUA2 and do the PWWP domains of the HUA2 and HULK proteins directly interact with chromatin at *FLC* and other loci?
3) Does disruption of the RPR domain affect the levels of *FLC* antisense and primary transcripts? Are there other genes whose sense and antisense transcript levels are affected by disruption of the RPR domains within HUA2 and the HULK proteins?
4) Does the disruption of the RPR domain in HUA2 and HULKs affect their interaction with AtCTD *in planta*?
5) What are the biochemical functions of HUA2 and HULK proteins?

References

Albà, M.M. and Pagès, M. (1998). Plant proteins containing the RNA- recognition motif. **3**.

Allison, L.A. et al. (1988). The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function. *Molecular and Cellular Biology***8**: 321–329.

Alvarez-Venegas, R. and Avramova, Z. (2012). Evolution of the PWWP-domain encoding genes in the plant and animal lineages. *BMC evolutionary biology***12**: 101.

Amasino, R. (2010). Seasonal and developmental timing of flowering. *The Plant journal : for cell and molecular biology***61**: 1001–13.

Amrani, N. et al. (1997). PCF11 encodes a third protein component of yeast cleavage and polyadenylation factor I. *Molecular and cellular biology***17**: 1102–9.

Anantharaman, V. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Research***30**: 1427–1464.

Andersson, C.R. et al. (2008). The FLX gene of Arabidopsis is required for FRI-dependent activation of FLC expression. *Plant cell physiology***49**: 191–200.

Arigo, J.T. et al. (2006). Regulation of yeast NRD1 expression by premature transcription termination. *Molecular cell***21**: 641–51.

Barillà, D. et al. (2001). Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America***98**: 445–50.

Bartkowiak, B. et al. (2011). Updating the CTD Story: From Tail to Epic. *Genetics Research International***2011**: 1–16.

Bedford, M.T. et al. (1997). FBP WW domains and the Abl SH3 domain bind to a specific class of proline-rich ligands. *the The European Molecular Biology Organization Journal***16**: 2376–2383.

Belostotsky, D. a and Rose, A.B. (2005). Plant gene expression in the age of systems biology: integrating transcriptional and post-transcriptional events. *Trends in plant science***10**: 347–53.

Blazquez, M. et al. (1998). Gibberellins promote flowering of arabidopsis by activating the LEAFY promoter. *The Plant cell***10**: 791–800.

Blázquez, M.A. and Weigel, D. (2000). Integration of floral inductive signals in Arabidopsis. *Nature***404**: 889–92.

Bonifaci, N. et al. (1997). Karyopherin beta2 mediates nuclear import of a mRNA binding protein. *Proceedings of the National Academy of Sciences***94**: 5055–5060.

Boulikas, T. (1993). Nuclear localization signals (NLS). *Critical reviews in eukaryotic gene expression***3**: 193–227.

Brameier, M. et al. (2007). NucPred--predicting nuclear localization of proteins. *Bioinformatics (Oxford, England)***23**: 1159–60.

Brown, J.W.S. and Simpson, C.G. (1998). SPLICE SITE SELECTION IN PLANT PRE-mRNA SPLICING. *Annual Review of Plant Physiology and Plant Molecular Biology***49**: 77–95.

Buratowski, S. (2005). Connections between mRNA 3' end processing and transcription termination. *Current opinion in cell biology***17**: 257–261.

Burn, J.E. et al. (1993). DNA methylation, vernalization, and the initiation of flowering. *Proceedings of the National Academy of Sciences of the United States of America***90**: 287–91.

Busch, M.A. et al. (1999). Activation of a Floral Homeotic Gene in Arabidopsis. *Science***285**: 585–587.

Campbell, M.A. et al. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC genomics***7**: 327.

Carty, S.M. and Greenleaf, A.L. (2002). Hyperphosphorylated C-terminal repeat domain-associating proteins in the nuclear proteome link transcription to DNA/chromatin modification and RNA processing. *Molecular cellular proteomics MCP***1**: 598–610.

Ceulemans, H. et al. (2000). The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? *FEBS Letters***456**: 1–5.

Chapman, R.D. et al. (2007). Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science***318**: 1780–1782.

Challa, Sathya Sheela (2009). Genetic and phenotypic analysis of the HUA2-LIKE (HULK) gene family in Arabidopsis thaliana. Ph.D.Thesis. The University of Western Ontario: Canada.

Chen, T. et al. (2004). The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. *Molecular and cellular biology***24**: 9048–58.

Chen, X. and Meyerowitz, E.M. (1999). HUA1 and HUA2 are two members of the floral homeotic AGAMOUS pathway. *Molecular Cell***3**: 349–360.

Cheng, Y. et al. (2003). Two RNA binding proteins, HEN4 and HUA1, act in the processing of AGAMOUS pre-mRNA in Arabidopsis thaliana. *Developmental Cell***4**: 53–66.

Choi, K. et al. (2011). The FRIGIDA complex activates transcription of FLC, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors. *The Plant Cell***23**: 289–303.

Clarke, J.H. and Dean, C. (1994). Mapping FRI, a locus controlling flowering time and vernalization response in Arabidopsis thaliana. *Molecular & general genetics : MGG***242**: 81–9.

Coen, E.S. and Meyerowitz, E.M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature***353**: 31–7.

Cokol, M. et al. (2000). Finding nuclear localization signals. *EMBO reports***1**: 411–5.

Conti, E. et al. (1998). Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. *Cell***94**: 193–204.

Corbesier, L. et al. (2007). FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science (New York, N.Y.)***316**: 1030–3.

Corden, J.L. (1990). Tails of RNA polymerase II. *Trends in Biochemical Sciences* **15**: 383–387.

Crevillén, P. and Dean, C. (2011). Regulation of the floral repressor gene FLC: the complexity of transcription in a chromatin context. *Current Opinion in Plant Biology***14**: 38–44.

Dahmus, M.E. (1994). The role of multisite phosphorylation in the regulation of RNA polymerase II activity. *Progress in nucleic acid research and molecular biology***48**: 143–79.

Dean, C. (1993). Advantages of Arabidopsis for Cloning Plant Genes. *Philosophical Transactions of the Royal Society B: Biological Sciences***342**: 189–195.

Dhayalan, A. et al. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *The Journal of biological chemistry***285**: 26114–20.

Dichtl, B. et al. (2002). Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *the The European Molecular Biology Organization Journal***21**: 4125–4135.

Doerks, T. (2002). Systematic Identification of Novel Protein Domain Families Associated with Nuclear Functions. *Genome Research***12**: 47–56.

Doyle, M.R. et al. (2005). HUA2 is required for the expression of floral repressors in Arabidopsis thaliana. *The Plant Journal***41**: 376–385.

Egloff, S. et al. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science***318**: 1777–1779.

de Felippes, F.F. and Weigel, D. (2010). Transient assays for the analysis of miRNA processing and function. *Methods in molecular biology (Clifton, N.J.)***592**: 255–64.

Ge, Y.-Z. et al. (2004). Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *The Journal of biological chemistry***279**: 25447–54.

Giakountis, A. and Coupland, G. (2008). Phloem transport of flowering signals. *Current opinion in plant biology***11**: 687–94.

Goodall, G.J. and Filipowicz, W. (1989). The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell***58**: 473–483.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. *Rna New York Ny***6**: 1197–1211.

He, Y. (2012). Chromatin regulation of flowering. *Trends in plant science*.

He, Y. (2009). Control of the transition to flowering by chromatin modifications. *Molecular plant***2**: 554–564.

He, Y. et al. (2004). PAF1-complex-mediated histone methylation of FLOWERING LOCUS C chromatin is required for the vernalization-responsive, winter-annual habit in Arabidopsis. *Genes & Development***18**: 2774–2784.

Hengartner, C.J. et al. (1998). Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Molecular Cell***2**: 43–53.

Henikoff, S. and Shilatifard, A. (2011). Histone modification: cause or cog? *Trends in Genetics***27**: 389–396.

Hirose, Y. and Manley, J.L. (2000). RNA polymerase II and the integration of nuclear events RNA polymerase II and the integration of nuclear events. *Genes Development***14**: 1415–1429.

Hornyik, C. et al. (2010). The spen family protein FPA controls alternative cleavage and polyadenylation of RNA. *Developmental Cell***18**: 203–213.

Hsieh, J.C. et al. (1998). Novel nuclear localization signal between the two DNA-binding zinc fingers in the human vitamin D receptor. *Journal of cellular biochemistry***70**: 94–109.

Huijser, P. and Schmid, M. (2011). The control of developmental phase transitions in plants. *Development (Cambridge, England)***138**: 4117–29.

Irie, Y. et al. (2000). Molecular cloning and characterization of Amida, a novel protein which interacts with a neuron-specific immediate early gene product arc, contains novel nuclear localization signals, and causes cell death in cultured cells. *The Journal of biological chemistry***275**: 2647–53.

J McCullough, A. and A Schuler, M. (1997). Intronic and exonic sequences modulate 5’ splice site selection in plant nuclei. *Nucleic Acids Research***25**: 1071–1077.

Jack, T. (2002). New members of the floral organ identity AGAMOUS pathway. *Trends in Plant Science***7**: 286–287.

Jaeger, K.E. et al. (2006). The control of flowering in time and space. *Journal of experimental botany***57**: 3415–8.

Jaeger, K.E. and Wigge, P.A. (2007). FT protein acts as a long-range signal in Arabidopsis. *Current biology : CB***17**: 1050–4.

Jiang, D. et al. (2009). Establishment of the winter-annual growth habit via FRIGIDA-mediated histone methylation at FLOWERING LOCUS C in Arabidopsis. *The Plant cell***21**: 1733–46.

Johanson, U. et al. (2000). Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. *Science (New York, N.Y.)***290**: 344–7.

KAY, B.K. et al. (2000). The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J***14**: 231–241.

Kang, C.H. et al. (2009). Arabidopsis thaliana PRP40s are RNA polymerase II C-terminal domain-associating proteins. *Archives of biochemistry and biophysics***484**: 30–8.

Karimi, M. et al. (2002). GATEWAY vectors for Agrobacterium-mediated plant transformation. *Trends in Plant Science***7**: 193–195.

Ko, C.H. et al. (1998). U-richness is a defining feature of plant introns and may function as an intron recognition signal in maize. *Plant Molecular Biology***36**: 573–583.

Ko, J.-H. et al. (2010a). Growth habit determination by the balance of histone methylation activities in Arabidopsis. *the The European Molecular Biology Organization Journal***29**: 3208–3215.

Ko, J.-H. et al. (2010b). Growth habit determination by the balance of histone methylation activities in Arabidopsis. *The EMBO journal***29**: 3208–15.

Kobayashi, Y. and Weigel, D. (2007). Move on up, it's time for change--mobile signals controlling photoperiod-dependent flowering. *Genes & development***21**: 2371–84.

Komarnitsky, P. et al. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Development***14**: 2452–2460.

Komiya, R. et al. (2009). A gene network for long-day flowering activates RFT1 encoding a mobile flowering signal in rice. *Development (Cambridge, England)***136**: 3443–50.

Koornneef, M. et al. (1998). GENETIC CONTROL OF FLOWERING TIME IN ARABIDOPSIS. *Annual review of plant physiology and plant molecular biology***49**: 345–370.

Lambermon, M.H. et al. (2000). UBP1, a novel hnRNP-like protein that functions at multiple steps of higher plant nuclear pre-mRNA maturation. *The EMBO journal***19**: 1638–49.

Lambermon, M.H.L. et al. (2002). UBA1 and UBA2 , Two Proteins That Interact with UBP1 , a Multifunctional Effector of Pre-mRNA Maturation in Plants UBA1 and UBA2 , Two Proteins That Interact with UBP1 , a Multifunctional Effector of Pre-mRNA Maturation in Plants.

Lee, I. et al. (1993). Analysis of naturally occurring late flowering in Arabidopsis thaliana. *Molecular general genetics MGG***237**: 171–176.

Lee, I. et al. (1994). The late-flowering phenotype of FRIGIDA and mutations in LUMINIDEPENDENS is suppressed in the Landsberg erecta strain of Arabidopsis. *The Plant Journal***6**: 903–909.

Lee, J.H. et al. (2007). Role of SVP in the control of flowering time by ambient temperature in Arabidopsis. *Genes & development***21**: 397–402.

Lenhard, M. et al. (2001). Termination of stem cell maintenance in Arabidopsis floral meristems by interactions between WUSCHEL and AGAMOUS. *Cell***105**: 805–814.

Li, D. et al. (2008). A repressor complex governs the integration of flowering signals in Arabidopsis. *Developmental cell***15**: 110–20.

Li, J. et al. (2001). HUA1, a regulator of stamen and carpel identities in Arabidopsis, codes for a nuclear RNA binding protein. *The Plant Cell***13**: 2269–2282.

Lifschitz, E. et al. (2006). The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. *Proceedings of the National Academy of Sciences of the United States of America***103**: 6398–403.

Liu, F. et al. (2010). Targeted 3' processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing. *Science (New York, N.Y.)***327**: 94–7.

Liu, F. et al. (2007). The Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC. *Molecular Cell***28**: 398–407.

Lohmann, J.U. et al. (2001). A molecular link between stem cell regulation and floral patterning in Arabidopsis. *Cell***105**: 793–803.

Lorković, Z.J., Wieczorek Kirk, D. a, Lambermon, M.H., et al. (2000). Pre-mRNA splicing in higher plants. *Trends in plant science***5**: 160–7.

Lorković, Z.J., Wieczorek Kirk, D.A., Klahre, U., et al. (2000). RBP45 and RBP47, two oligouridylate-specific hnRNP-like proteins interacting with poly(A)+ RNA in nuclei of plant cells. *RNA (New York, N.Y.)***6**: 1610–24.

Macknight, R. et al. (1997). FCA, a Gene Controlling Flowering Time in Arabidopsis, Encodes a Protein Containing RNA-Binding Domains. *Cell***89**: 737–745.

Mathieu, J. et al. (2007). Export of FT protein from phloem companion cells is sufficient for floral induction in Arabidopsis. *Current biology : CB***17**: 1055–60.

Maurer-Stroh, S. et al. (2003). The Tudor domain "Royal Family": Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends in Biochemical Sciences* **28**: 69–74.

Meinke, D.W. et al. (1998). Arabidopsis thaliana: a model plant for genome analysis. *Science (New York, N.Y.)***282**: 662, 679–82.

Merritt, H. et al. (1997). Internal AU-rich elements modulate activity of two competing 3' splice sites in plant nuclei. *The Plant Journal***12**: 937–943.

Meyerowitz, E.M. and Somerville, C.. (1994). Arabidopsis (New York: Cold Spring Harbor Laboratory Press).

Michaels, S.D. et al. (2003). Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behavior in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America***100**: 10102–7.

Michaels, S.D. (1999). FLOWERING LOCUS C Encodes a Novel MADS Domain Protein That Acts as a Repressor of Flowering. *THE PLANT CELL ONLINE***11**: 949–956.

Michaels, S.D. et al. (2004). FRIGIDA-related genes are required for the winter-annual habit in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America***101**: 3281–3285.

Michaels, S.D. and Amasino, R.M. (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *The Plant Cell***11**: 949–956.

Michaels, S.D. and Amasino, R.M. (2001). Loss of FLOWERING LOCUS C Activity Eliminates the Late-Flowering Phenotype of FRIGIDA and Autonomous Pathway Mutations but Not Responsiveness to Vernalization. *The Plant Cell***13**: 935–941.

Moede, T. et al. (1999). Identification of a nuclear localization signal, RRMKWKK, in the homeodomain transcription factor PDX-1. *FEBS Letters***461**: 229–234.

Morris, D.P. and Greenleaf, A.L. (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *The Journal of Biological Chemistry***275**: 39935–39943.

Mouradov, A. et al. (2002). Control of flowering time: interacting pathways as a basis for diversity. *The Plant Cell***14 Suppl**: S111–S130.

Möller-Steinbach, Y. et al. (2010). Flowering time control. *Methods in molecular biology (Clifton, N.J.)***655**: 229–37.

Napp-Zinn, K. (1957). No TitleUntersuchungen zur genetik des kaltebedurfnisses bei Arabidopsis thaliana (L.) Heynh. Z Indukt Abstammungs- Vererbungsl. **88**: 253–285.

Noble, C.G. et al. (2005). Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nature structural & molecular biology***12**: 144–51.

Orphanides, G. and Reinberg, D. (2002). A Unified Theory of Gene Expression Review. *Cell***108**: 439–451.

Park, J.-I. et al. (2006). Molecular characterization of two anther-specific genes encoding putative RNA-binding proteins, AtRBP45s, in Arabidopsis thaliana. *Genes & genetic systems***81**: 355–9.

Patturajan, M. et al. (1998). A Nuclear Matrix Protein Interacts with the Phosphorylated C-Terminal Domain of RNA Polymerase II. *Molecular and Cellular Biology***18**: 2406–2415.

Peal, L. et al. (2011). Phylogenetic and expression analysis of RNA-binding proteins with triple RNA recognition motifs in plants. *Molecules and cells***31**: 55–64.

Phatnani, H.P. and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes & Development***20**: 2922–2936.

Poduska, B. et al. (2003). The synergistic activation of FLOWERING LOCUS C by FRIGIDA and a new flowering gene AERIAL ROSETTE 1 underlies a novel morphology in Arabidopsis. *Genetics***163**: 1457–65.

Pyke, K. (1994). Tansley Review No. 75 Arabidopsis - its use in the genetic and molecular analysis of plant morphogenesis. *New Phytologist***128**: 19–37.

Qiu, C. et al. (2002). The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nature Structural Biology***9**: 217–224.

Quesada, V. et al. (2003). Autoregulation of FCA pre-mRNA processing controls Arabidopsis flowering time. *the The European Molecular Biology Organization Journal***22**: 3142–3152.

Reddy, A.S.N. (2001). Nuclear Pre-mRNA Splicing in Plants. *Critical Reviews in Plant Sciences***20**: 523–571.

Reddy, A.S.N. (2004). Plant serine/arginine-rich proteins and their role in pre-mRNA splicing. *Trends in Plant Science***9**: 541–547.

Rodriguez, C.R. et al. (2000). Kin28, the TFIIH-Associated Carboxy-Terminal Domain Kinase, Facilitates the Recruitment of mRNA Processing Machinery to RNA Polymerase II. *Molecular and Cellular Biology***20**: 104–112.

Rouse, D.T. et al. (2002). FLC, a repressor of flowering, is regulated by genes in different inductive pathways. *The Plant Journal***29**: 183–191.

SAS Institute Inc. 2003. SAS version 9.2. Cary NC.

Sadowski, M. et al. (2003). Independent functions of yeast Pcf11p in pre-mRNA 3′ end processing and in transcription termination. *The EMBO journal***22**: 2167–2177.

Sajja, Uday Bhaskar. (2009). Molecular characterization of Arabidopsis flowering regulator HUA2, a putative pre mRNA processing factor. Ph.D. Thesis. The University of Western Ontario: Canada.

Sanda, S.L. and Amasino, R.M. (1996). Interaction of FLC and late-flowering mutations in Arabidopsis thaliana. *Molecular & general genetics : MGG***251**: 69–74.

Saunders, A. et al. (2006). Breaking barriers to transcription elongation. *Nature reviews Molecular cell biology***7**: 557–567.

Schmidt, F. et al. (2010). A proteomic analysis of oligo(dT)-bound mRNP containing oxidative stress-induced Arabidopsis thaliana RNA-binding proteins ATGRP7 and ATGRP8. *Molecular biology reports***37**: 839–45.

Shaner, N.C. et al. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature biotechnology***22**: 1567–1572.

Sheldon, C.C. (1999). The FLF MADS Box Gene: A Repressor of Flowering in Arabidopsis Regulated by Vernalization and Methylation. *THE PLANT CELL ONLINE***11**: 445–458.

Simpson, G.G. and Dean, C. (2002). Arabidopsis, the Rosetta stone of flowering time? *Science***296**: 285–9.

Siomi, H. and Dreyfuss, G. (1997). RNA-binding proteins as regulators of gene expression. *Current opinion in genetics & development***7**: 345–53.

Smyth, D.R. et al. (1990). Early flower development in Arabidopsis. *The Plant cell***2**: 755–67.

Steeves, T.A. and Sussex, I.M. (1989). Patterns in Plant Development (Cambridge University Press: Cambridge).

Steinmetz, E.J. et al. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature***413**: 327–31.

Steinmetz, E.J. and Brow, D.A. (1998). Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. *Proceedings of the National Academy of Sciences of the United States of America***95**: 6699–6704.

Tamaki, S. et al. (2007). Hd3a protein is a mobile flowering signal in rice. *Science (New York, N.Y.)***316**: 1033–6.

Tietjen, J.R. et al. (2010). Chemical-genomic dissection of the CTD code. *Nature structural & molecular biology***17**: 1154–61.

Tinland, B. et al. (1992). The T-DNA-linked VirD2 protein contains two distinct functional nuclear localization signals. *Proceedings of the National Academy of Sciences of the United States of America***89**: 7442–7446.

Truant, R. and Cullen, B.R. (1999). The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals. *Molecular and cellular biology***19**: 1210–7.

Ursic, D. et al. (2008). Detecting phosphorylation-dependent interactions with the C-terminal domain of RNA polymerase II subunit rpb1p using a yeast two-hybrid assay. *RNA Biology***5**: 1–4.

Vezzoli, A. et al. (2010). Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nature structural & molecular biology***17**: 617–9.

Voinnet, O. et al. (2000). A viral movement protein prevents spread of the gene silencing signal in Nicotiana benthamiana. *Cell***103**: 157–167.

Wang, B.-B. and Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. *Proceedings of the National Academy of Sciences of the United States of America***103**: 7175–7180.

Wang, B.-B. and Brendel, V. (2004). The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. *Genome biology***5**: R102.

Wang, Q. et al. (2007). HUA2 caused natural variation in shoot morphology of A. thaliana. *Current Biology***17**: 1513–1519.

Wang, Y. et al. (2009). Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. *Molecular Cell***33**: 428–437.

Weigel, D. and Meyerowitz, E.M. (1994). The ABCs of floral homeotic genes. *Cell***78**: 203–9.

Williamson, M.P. (1994). The structure and function of proline-rich regions in proteins. *The Biochemical journal***297**: 249–260.

Willis, C.G. et al. (2008). Phylogenetic patterns of species loss in Thoreau's woods are driven by climate change. *Proceedings of the National Academy of Sciences of the United States of America***105**: 17029–33.

Wu, H. et al. (2011). Structural and histone binding ability characterizations of human PWWP domains. *PloS one***6**: e18919.

Xu, L. et al. (2008). Di- and Tri- but Not Monomethylation on Histone H3 Lysine 36 Marks Active Transcription of Genes Involved in Flowering Time Regulation and Other Processes in Arabidopsis thaliana. *Molecular and Cellular Biology***28**: 1348–1360.

Yang, B. et al. (2003). Potent suppression of viral infectivity by the peptides that inhibit multimerization of human immunodeficiency virus type 1 (HIV-1) Vif proteins. *The Journal of biological chemistry***278**: 6596–602.

Yant, L. et al. (2010). Orchestration of the floral transition and floral development in Arabidopsis by the bifunctional transcription factor APETALA2. *The Plant cell***22**: 2156–70.

Yu, X. and Michaels, S.D. (2010). The Arabidopsis Paf1c Complex Component CDC73 Participates in the Modification of FLOWERING LOCUS C Chromatin. *Plant Physiology***153**: 1074–1084.

Yuryev, A. et al. (1996). The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proceedings of the National Academy of Sciences of the United States of America***93**: 6975–80.

Zhao Mingyu, Zhang Zhengbin, Chen Shouyi, Z.J. and S.H. (2012). WRKY transcription factor superfamily: Structure, origin and functions. *African Journal of Biotechnology***11**: 8051–8059.

Appendix I

Combination of fusion proteins used for yeast two-hybrid assay

Table A1. List of GAL4 AD fusion proteins co-expressed with HUA2 GAL4 DB in yeast cells to confirm interaction observed in yeast two-hybrid screen.

GAL4 DB	GAL4 AD
HUA2	FCA
HUA2	PRP40(ww)
HUA2	RBP45
HUA2	UBP1

Table A2. List of GAL4 AD of fusion proteins co-expressed with different *HUA2* alleles fused withGAL4 DB in yeast cells to check which domain in HUA2 is responsible for interaction

GAL4 DB	GAL4 AD			
HUA2	FCA	PRP40(ww)	RBP45	UBP1
hua2-1	FCA	PRP40(ww)	RBP45	UBP1
hua2-5	FCA	PRP40(ww)	RBP45	UBP1
CT-HUA2	FCA	PRP40(ww)	RBP45	UBP1

Table A3. List of GAL4 AD fusion proteins co expressed with CT-HULK1-GAL4 DB and CT-HULK3 GAL4 DB in yeast cells

GAL4 DB	GAL4 AD			
CT-HULK1	FCA	PRP40(ww)	RBP45	UBP1
CT-HULK3	FCA	PRP40(ww)	RBP45	UBP1

Table A4. List of GAL4 AD fusion proteins co expressed with HUA2-GAL4 DB and CT-HUA2 GAL4 DB in yeast cells to check for homo/hetrodimerization of HUA2 protein family

GAL4 DB	GAL4AD			
HUA2	HUA2	HULK1	HULK2	HULK3
CT-HU2	HUA3	HULK1	HULK2	HULK3

Table A5. List of GAL4 AD fusion proteins co expressed with CT-HULK1-GAL4 DB and CT-HULK3 GAL4 DB in yeast cells to check for homo/hetrodimerization of HUA2 protein family

GAL4 DB	GAL4 AD
CT-HULK1	HUA2
CT-HULK3	HUA2

Table A6. List of GAL4 AD fusion proteins co expressed with AtCTD-GAL4 DB in yeast cells to check AtCTD's interaction with HUA2 and HUA2ΔRPR

GAL4 DB	GAL4 AD
AtCTD	PRP40
AtCTD	HUA2
AtCTD	HUA2-ΔRPR

Table A7. List of GAL4 AD fusion proteins co expressed with HUA2-GAL4 DB and HUA2-ΔRPR GAL4 DB in yeast cells to check if HUA2 and HUA2ΔRPR interact with AtCTD

GAL4 DB	GAL4 AD
HUA2	PRP40
HUA2	AtCTD
CT-HUA2	AtCTD
HUA2-ΔRPR	AtCTD

Table A8. List of GAL4 AD fusion proteins co expressed with AtCTD-GAL4 DB in yeast cells to check interaction between AtCTD and the HULKs

GAL4 DB	GAL4 AD
AtCTD	HULK1
AtCTD	HULK2
AtCTD	HULK3

Table B1. List of GAL4 AD fusion proteins co expressed with HUA2-GAL4 DB and PCF11-GAL4 DB in yeast cells to check HUA2's interaction with different phosphorylation states of yeast CTD

GAL4 DB	GAL4 AD
PCF11	AtCTD
HUA2	A2S5
HUA2	S2S5
HUA2	S2A5

Curriculum Vitae

Name:	Preetam Janakirama
Post-secondary Education and Degrees:	Bangalore University St Joseph's College of Arts and Science, Bangalore, India 1999-2002 B.Sc. Bangalore University CMRIMS, Bangalore, India 2002-2004 M.Sc. The University of Western Ontario London, Ontario, Canada 2007-2013 Ph.D.
Honours and Awards:	Western Graduate Research Scholarship 2007-2011 The Ontario/Baden-Württemberg Summer Research Scholarship 2010
Related Work Experience	Research Assistant National Center for Biological Science, Bangalore 2004-2005 Research Assistant St John's research institute of population health and clinical research, Bangalore, India 2005 Research Assistant Indian Institute of Science, Bangalore, India 2005-2006 Teaching Assistant The University of Western Ontario 2007-2012